Ethical Issues for AI-Solutions in Business: Hype or..?

Conference Paper · February 2023

CITATIONS
0

READS
13

1 author:

Sergei Kladko
Innopolis University
12 PUBLICATIONS **0** CITATIONS

SEE PROFILE

Ethical Issues for AI-Solutions in Business: Hype or..?

Sergei Kladko, Innopolis University (s2kladko@gmail.com)

Abstract:
The abstract considers the issues of ethics in the solutions on the basis of artificial intelligence (AI) with regards to modern business environment. The author believes that the so-called 'ethical' challenges that find their practical implementation in AI solutions in various types of discrimination have already become a serious threat both for the IT-developers and for the companies that apply such tools in their everyday business practice. The recent years have witnessed an unprecedented growth in the number of cases when the ignorance to the ethical bias in AI solutions resulted in financial and reputational damage. To help business community overcome such challenges, the author, on the basis of the analysis of the corpus of practical, provides three suggestions regarding more active introduction of ethics in AI solutions and, on the other hand, avoiding the so-called 'ethics washing' as a potentially dangerous practice which impedes technological and business development.

Keywords: *AI ethics, artificial intelligence, technologies in business, business ethics*

1. AI Market: Lucrative Prospects and Expensive Biases

Today, the global market of artificial intelligence (AI), with a projected growth from $432 billion in 2022 to $900 billion by 2026 [International Data Corporation 2022], cannot but amaze with its prospects in the near future. At the same time in the Russian Federation the AI market volume is predicted to amount to 555.1 million dollars by 2024 [TAdviser 2022] (however, other forecasts look even more optimistic, assuming the market growth of up to 160 billion rubles [Ministry of Digital Development, 2019]). Correspondingly, the companies increase their expenditures on the implementation of AI solutions with the hope to gain a stable financial position in the future and, most importantly, remain competitive in their markets. For instance, according to the survey in the United States, more than a third of high-margin companies spends on AI from $51 to 100 million, and seven out of ten organizations spend $1 million or more of their budget on AI [Venturebeat 2022] (and this does not include smaller companies where spending on AI solutions, in most cases, exceed $50 thousand per year).

There is no need to say that the abovementioned figures and prospects for the development of the AI solutions market are staggering. However, today various actors (gradually, but quite confidently) interacting within the AI ecumene are asking questions related not only to the financial prospects for the implementation of these solutions but also to the ways of how to make AI solutions more 'responsible' in terms of the moral and ethical attitudes of the relevant target audiences. One can, of course, argue that ethical issues in business have always been considered a kind of semantic load, which, under favorable circumstances, could also be monetized. However, the current practice demonstrates that underestimating (or completely ignoring) the ethical issues that various TA's care about when interacting with AI solutions in education, medicine, litigation, or recruting leads not only to reputational, but also to significant financial costs. This can be seen in the results of a joint survey by DataRobot and the World Economic Forum, in which the participants declare that the presence of biases in data has already cost them the loss of income (62% of respondents), customers (61%) or employees (43%), and some of them (35%) suffered serious financial costs caused by lawsuits [DataRobot 2022]. Moreover, 6% of respondents were forced to admit that underestimating data biases in AI solutions led to significant damage to the company's brand, the biggest cost for the business today. On the other hand, behind the dry descriptions of cases of cultural and ethical failures of AI algorithms are dozens of real people who have lost their jobs, their good name, and sometimes their lives. The

resulting multimillion-dollar lawsuits and reputational losses, directly related to the inclusion of data filled with racial, national, gender and cultural biases in AI algorithms, clearly do not allow businesses to regard this problem as of secondary importance and not related to the financial stability of companies.

2. Key Challenges to Ethics in AI Solutions

First of all, it is important to emphasize that, for this paper, AI ethical issues are considered in regards of the so-called narrow (weak) AI characterized by the limited use of data, depending on who created or processed it. Hence, one of the main problems in the development and further operation of an "ethically correct" AI solution is the significant subjectivism of the concepts of "morality". At the same time, as researchers today admit, it is impossible for any developer to get rid totally of their own cultural stereotypes when creating an appropriate AI solution, which, as a result, will unambiguously contain a set of cultural associations and prejudices of actors (developers primarily) [Caliskan, 2017, p. 184]. At the same time, the norms and values of the respective TA's will not always coincide with the cultural characteristics of the developers, and, in some cases, seriously contradict them. Accordingly, it is not possible to talk about the possibility of a unanimous opinion on the same concepts of "good" and "evil". It is not surprising, therefore, that all this causes a natural negative reaction both from developers who are being blamed for the "unethical" behavior of AI algorithms and from the business community, which is trying to find the right balance between community expectations and preventing the rise in the cost of innovative solutions. The global survey conducted by the Pew Research Center puts the concern about the generally accepted terminology of AI-ethics at the top of the list of the main problems in creating AI solutions that meet the ethical expectations of various actors. Other anxieties voiced by respondents also cast doubt on whether ethics may soon become an important issue when working on AI solutions. First of all, the main players in the market (large corporations and governments in general) are not very concerned with such issues themselves, preferring other criteria for evaluating the effectiveness and importance of the corresponding product or service. Finally, developers and businesses are absolutely right by stating that an excessive focus on resolving ethical issues in AI solutions will lead to a serious increase in the cost of the latter (even collecting the most diverse data means additional costs) and, as a result, to the final defeat in the technological race. It should be recognized that such fears are clearly not unfounded, and, as a result, the lack of clear answers to the questions above means that ethical issues continue to be sacrificed in the process of constant competition in the technological market. This is exactly the answer to the question of whether most AI systems will be based on ethical principles by 2030 by almost 68% of practitioners (including the heads of the largest IT companies) within the above question [Pew Research Center, 2021].

Finally, even the creation of an AI solution that could turn out to be as ethically correct as possible may not mean that its creators will not receive the same reproaches after some time as the developers who ignored the moral aspect in data processing. Today there are serious concern about the so-called 'ethics greenwashing' which can replace the process of searching for the balance between ethical expectations and the real-life situation, thus discrediting totally the very idea of AI-ethics. An illustration of this is Microsoft's attempt to fix the epic fail with the Tay chatbot by launching the Zo chatbot, an important characteristic of which was maximum neutrality in relation to the most sensitive topics for users (religion, race, etc.), while all concepts associated with individual cultures (from their history to positive judgments about them) fell under the ban. As a result, the new chatbot was accused of censorship, narrow-mindedness, and… insufficiently ethical behavior [Stuart-Ulin 2018 (2022)].

3. Case Study and Conclusions

In an attempt to develop recommendations for a more balanced consideration of ethical factors in the development of AI solutions, the author analyzed a corpus of 24 practical cases collected from open sources over the period from 2017 to 2022. Cases were divided into groups in the following areas: medicine, law and litigation, education, and recruitment. The development of

each case was tracked in open sources (including lessons learned from it for other companies), on the basis of which a conclusion was made about what has changed for a particular area in terms of the "ethical" implementation of AI solutions over a five-year period. All this allowed the author to come to the following conclusions:

1) In the field of healthcare, the number of cases of unethical interpretation of data by AI algorithms remains quite high; however, testing has become more thorough in recent years (which, given the scope, is not surprising). Prejudices are connected, first of all, with the insufficient representativeness of the data in relation to the racial / national or gender affiliation of the person being tested. At the same time, medicine remains the only area where it is virtually impossible to find in the public domain the name of the project or the company that made such mistakes (there are only general description of cases).

2) Over the past three years, the number of cases of “unethical” use and interpretation of data in AI solutions in the field of education has increased significantly. In many ways, this was due to the COVID-19 pandemic, when, in order to save the educational process, it was sometimes necessary to put into practice not fully tested technologies. As a result, racial and social biases embedded in AI solutions have had a negative impact on the results of the most important assessment exams in the US [Reeves 2021] and the UK [Coughlan 2020], or, for example, when objectively considering a low GPA for university admission [Burke 2020]. Moreover, there was a case when an ethical bias in terms of income or place of study affected several thousand people in various countries (certification under the International Baccalaureate program in 2020 [Evgeniou 2020]). At the moment, the penetration of AI technologies in education continues to grow dynamically, which is likely to lead to the emergence of a considerable number of cases when the ethical component leads to significant conflicts.

3) Within the realm of litigation and law enforcement, after the COMPASS and HART scandals, discriminatory prospects based on the use of biased data seemed to be significantly reduced. However, a new scandal soon followed with the PredPol system (2018-2021), which continued the unethical activities of its predecessors, thus raising again the question of the correctness of using of such solutions in the field of law enforcement [Sankin 2021]. However, the developers of some new systems claim that they were able to take into account previous mistakes and that their systems have passed (and are passing) the most serious ethics test (although even in this case there are those who seriously question such claims) [Simonite T. 2019].

4) The field of recruitment continues to present a fairly serious ethical challenge for AI solutions. The history of Amazon's AI recruiting tools, accused of gender discrimination, has had little effect on the positive dynamics of the emergence of new AI solutions in this area. To date, applicants continue to complain about gender and age discrimination, as well as about ableism. However, the developers of such solutions tend to respond rather quickly to criticism regarding the ethical inconsistency of their products, which in fact strengthens the reputation of the brand.

4. Solutions and Proposals

In an attempt to resolve the issues about how AI solutions can best meet the ethical expectations of their respective target TA’s, experts (both in business and in academia) offer various solutions. In relation to the distant future of AI products, it seems possible to talk about a certain convergence of human and machine behavior, which is already being recorded today by some scientists who note that a person surrounded by mechanisms imitates the behavior of electronic machines which, in their turn, acquire more and more human characteristics [Mazzara 2021]. Moreover, there is also the possibility that AI-based systems are beginning (albeit on a short-term basis) to influence the wider human culture [Brinkmann 2022], which, in principle, could resolve many of the current ethical discrepancies. However, all these transhumanist ideas do not solve the current problems that continue to negatively affect the development of AI solutions in industry and service.

First of all, it seems necessary to assist businesses and developers in developing methodological solutions for the formation of datasets with diverse and maximally free from ethical prejudices data. To date, the author is a member of a group that develops such methods by order of enterprises. These instructions contain not only recommendations on creating such datasets, but also a selection of practical cases, on the basis of which business customers can form a clear understanding of what a specific bias in a data set means and, most importantly, what the consequences of ignoring this problem.

Further, it can be interesting to develop the methods for a preliminary cultural and ethical audit of AI solutions. As part of these decisions, the author is participating in the working out of the foundations for the so-called cultural and ethical "red teaming". It should be noted that the classic red-teaming technique itself has long been very successfully used to find weaknesses in IT solutions, and it can be assumed that it can be reworked in order to identify weaknesses in forecasts and assumptions regarding the correct interpretation of the cultural and ethical intentions of customers and users. The main task of such teams (either formed within the company or a specially trained and specially certified third party) should be a simulated scheduled verification of an AI solution from the point of view of its most adequate subsequent normative-value perception by the relevant target audience.

Finally, it is important to intensify the creation of an institution of specialists in cultural and ethical facilitation in the process of implementing AI solutions. The presence of a trained certified expert mediator, who can help actors detect cultural and ethical contradictions when implementing a project and not allow ethical consideration transform into ‘ethics washing’, allows developers and customers not only to receive professional assistance, but also, most importantly, to significantly reduce the time and financial costs of independently conducting ongoing ethical AI audits for the solutions.

References

1. Brinkmann L., Gezerli D., Müller K., Rahwan I., Pescetelli N. 2022 Hybrid social learning in human-algorithm cultural transmission // Royal Society Publishing. May 25. https://doi.org/10.1098/rsta.2020.0426 (accessed September 12, 2022)
2. Burke L. The Death and Life of an Admissions Algorithm // Inside Higher Ed, December 14, 2020 https://www.insidehighered.com/admissions/article/2020/12/14/u-texas-will-stop-using-controversial-algorithm-evaluate-phd (Accessed November 07, 2021)
3. Caliskan, A, Bryson J., Narayanan A. Semantics derived automatically from language corpora contain human-like biases// Science, 356 (6334). 2017. C. 183-186.
4. Coughlan S. Why did the A-level algorithm say no? // BBC August 14, 2020. https://www.bbc.com/news/education-53787203 (Accessed November 11, 2021).
5. DataRobot’s State of AI Bias Report Reveals 81% of Technology Leaders Want Government Regulation of AI Bias //DataRobot January 18, 2022 https://www.datarobot.com/newsroom/press/datarobots-state-of-ai-bias-report-reveals-81-of-technology-leaders-want-government-regulation-of-ai-bias/ (accessed September 03, 2022)
6. Evgeniou T., Hardoo D., Ovchinnikov A. What Happens When AI is Used to Set Grades? //Harvard Business Review August 13, 2020 https://hbr.org/2020/08/what-happens-when-ai-is-used-to-set-grades (Accesses September 13, 2022).
7. International Data Corporation. 2022. IDC Forecasts 18.6% Compound Annual Growth for the Artificial Intelligence Market in 2022-2026 Report, July 29 https://www.idc.com/getdoc.jsp?containerId=prEUR249536522 (accessed August 30, 2022)
8. Mazzara M, Zhdanov P, Bahrami M.R. , Aslam H, Kotorov Iu, Imam M, and Salem H. Education after COVID-19 / In book: Smart and Sustainable Technology for Resilient Cities and Communities. Springer. June 2021. https://www.researchgate.net/publication/351587748_Education_after_COVID-19 (accessed August 20, 2022)

9. Ministry of Digital Development, Communications and Mass Media of the Russian Federation. 2019. Roadmap for the Development of the Digital Technology "Neurotechnologies and Artificial Intelligence" (in Russian), p.10 https://digital.gov.ru/ru/documents/6658/ (accessed August 30, 2022)
10. Pew Research Center. 2021. Experts Doubt Ethical AI Design Will Be Broadly Adopted as the Norm Within the Next Decade (Report). June 16. https://www.pewresearch.org/internet/2021/06/16/experts-doubt-ethical-ai-design-will-be-broadly-adopted-as-the-norm-within-the-next-decade/ (accessed August 01, 2022)
11. Sankin A., Mehrota D., Mattu S., Cameron D., Gilbertson A., Lempres D, Lash J. 2021. Crime Prediction Software Promised to Be Free of Biases. New Data Shows It Perpetuates Them. Gismodo, February 12. https://gizmodo.com/crime-prediction-software-promised-to-be-free-of-biases-1848138977 (accessed September 02. 2022)
12. Simonite T. 2019. The Best Algorithms Struggle to Recognize Black Faces Equally. Wired. July 22. https://www.wired.com/story/best-algorithms-struggle-recognize-black-faces-equally/ (accessed August 31, 2022)
13. Stuart-Ulin Chloe. 2018 (updated 2022). Microsoft's politically correct chatbot is even worse than its racist one. Quartz, July 21 (July 31). https://qz.com/1340990/microsofts-politically-correct-chat-bot-is-even-worse-than-its-racist-one/ (accessed August 13, 2022)
14. Richard V. Reeves and Dimitrios Halikias Race gaps in SAT scores highlight inequality and hinder upward mobility https://www.brookings.edu/research/race-gaps-in-sat-scores-highlight-inequality-and-hinder-upward-mobility/ (Accessed September 10, 2022).
15. TAdviser. 2022. Solutions on the Basis of Artificial Intelligence (in Russian). January 08 shorturl.at/ikuw9 (accessed September 04, 2022)
16. Venturebeat. 2022. Report: 70% of orgs are spending $1M or more on AI. February 22. https://venturebeat.com/ai/report-70-of-orgs-are-spending-1m-or-more-on-ai/ (accessed September 06, 2022)