B. Business Impact

AI, Friend or Foe? A Proposal for a Human-Centered Approach to AI Business Implications

ESCP Impact Paper No 2021-07-EN

Lorena Blasco-Arcas & Hsin-Hsuan Meg Lee
ESCP Business School

ESCP RESEARCH INSTITUTE OF MANAGEMENT (ERIM)

ESCP BUSINESS SCHOOL

ESCP Impact Paper No.2021-07-EN

AI, Friend or Foe? A Proposal for a Human-Centered Approach to AI Business Implications

Lorena Blasco-Arcas
ESCP Business School

Hsin-Hsuan Meg Lee
ESCP Business School

Abstract

Artificial Intelligence (AI) constitutes one of the most impactful developments in recent years for businesses and organisations in general. The pervasive use of data from voice assistants, visual data, visual recognition systems or internet-of-things applications, among others, provide AI systems with sources of data training to develop and improve their algorithms and consequently offer better-automated decision-making and predicting systems to companies in their internal processes.

However, this fast-paced and unstoppable trend raises many issues related to the technology, the actors involved, and the implementation and use of the outcomes. How organisations and actors in the AI ecosystem ensure that AI development is ethical and sustainable to the benefit of humans remains a primary concern moving forward. We propose to apply a human-centred view to the design of specific frameworks and regulatory systems.

Keywords: artificial intelligence, ethics, data, bias, discrimination.

AI, Friend or Foe? A Proposal for a Human-Centered Approach to AI Business Implications

Two Scottish people walked into an elevator that could only be activated by voice recognition. They struggled to make the elevator understand which floor they were going to. They then made their request with an American accent, hoping that it would be far easier to understand. This story is a comedy sketch by the Scottish Comedy Channel[1] a few years ago that has been widely shared online. The clip was hilarious because it was an experience shared by many, not just the Scottish. Sometimes when interacting with machines, smart machines in particular, people may feel dumbfounded, as if their intelligence were not on the same wavelength as the machines. While this is not the intention of AI development - to only interact with a few - for the "minorities" in the tech world, it certainly feels like it during user experiences.

The pervasive technological advances we have witnessed in recent years have generated unparalleled access to consumer data that is changing the competitive landscape and reshaping marketing activities. The global artificial intelligence (AI) software market is forecast to grow rapidly in the coming years, reaching around 126 billion U.S. dollars by 2025.[2] Artificial intelligence and many applications that utilise such technology have entered the marketplace. Along with growing concerns about the privacy issues related to data usage and data storage, there is also an increasing worry about the biases that exist in the algorithm development of all these tools. Biases in the technology world transcend spoken accents. Much like institutional racism, which requires fundamental shifts in the overall ecosystem, the problems in AI development also call for a similar change to create better output. To solve this issue, we propose to prioritise humans faced with technological advancement. In this paper, we first discuss the implications of AI in businesses and the challenges of current systematic biases, and finally propose a human-centred framework for the future roadmap.

The increasing application of AI in business

AI refers to "systems that display intelligent behaviour by analysing their environment and taking actions– with some degree of autonomy – to achieve specific goals." (European Commission, 2018). When classifying AI-based systems, we can distinguish between those that are purely software-based (e.g., Voice assistants, image analysis software, search engines, or face recognition systems) and those embedded into other hardware devices (e.g., robots, autonomous cars or any Internet-of-things application).

AI is considered one of the most popular technologies in business today due to three significant factors: the exponential growth of big data, access to cheap and scalable computational power, and the development of new and more sophisticated AI techniques (Overgood et al., 2019).

This development has been made possible by increasing digital data, available computing power and innovations coming from a few major players in the technology industry. According to Statista (2020), the volume of data created, captured, copied and consumed worldwide was around 59 zettabytes in 2020 and will reach 149 zettabytes by 2024.[3] This

[1] https://www.youtube.com/watch?v=HbDnxzrbxn4

[2] Global AI software market size 2018-2025. https://www.statista.com/statistics/607716/worldwide-artificial-intelligence-market-revenues/

[3] Volume of data/information created, captured, copied and consumed worldwide https://www.statista.com/statistics/871513/worldwide-data-created

increased access to large amounts of data creates massive commercial opportunities and poses new challenges for companies in processing, managing and computing the data while extracting value out of it (Erevelles et al., 2016; Sirivajah et al. 2017). The success of AI technology globally is forcing many existing companies to transform their business model and shift to AI.

A study of Fortune 1000 companies by New Vantage Partners has revealed that "almost 92% of businesses surveyed are increasing the pace of their big data and AI investments" in recent years. However, the collection and purchase of large datasets needed to train AI technology are not affordable to everyone. Consequently, only a few companies lead the machine learning and AI market (Soni et al., 2019).

The issue of increasing AI application in business is exacerbated as technology develops and access to computational power escalates too. Today, we have machine-learning algorithms, which might be trained through unsupervised learning, making decisions about credit, medical diagnoses, personalised recommendations, advertising and job opportunities, among other things. The decision-making is done, but in most cases it is a mystery how the final solution is achieved (Spice, 2016).

Data bias and discrimination practices in the AI field

Algorithmic bias is not a new phenomenon in AI. In the 1970s and 1980s, St. George's Hospital Medical School in the United Kingdom used a computer programme to do the initial screening of applicants. Based on a database of previous choices from admission staff, the program denied interviews to more than 60 applicants who were women and people with non-European names. The program used was not the problem itself but it perpetuated questionable racial and gender practices in the selection process (García, 2016).

To date, engineers have focused more on developing AI algorithms to solve complex problems than on the need to monitor and report their technology advances. "Today, the way a lot of AI is configured is basically as a black box (...) Neural networks are not good at explaining why they made a decision," explains Suranga Chandratillake, CEO of RhythmOne and General Partner at Balderton Capital. If people do not know how the technology came up with its results, they will not trust it.

We have already seen some examples of failures by the technology with the rise of discriminatory practices. For instance, in 2016, Microsoft released its self-learning chatbot Tay on Twitter. Supposed to be an experiment in "conversational understanding" able to learn language fundamentals over time, it could participate in a conversation by itself. However, it happened that the bot was developing racist and sexist traits on social media (The Verge, 2016). Another example occurred at MIT when Joy Buolamwini conducted a discriminatory experiment without knowing it while working on Facial Recognition. As a dark-skinned woman, she could not be recognised as precisely as her white friend by the AI. The results completely missed the point of the experience, and she found out that white women were recognised at 99% by the computer compared to 65% for black women. Was it a human intention to make the AI act like this? Maybe or maybe not. When this kind of power is given to the machine, the machine "encodes that bias" (Buolamwini, 2019). This does not mean that these artificial Intelligence tools were fundamentally flawed or designed to be racists.

Nevertheless, they were indeed biased by their designers and not controlled enough before going public. This is why data bias can lead to discrimination practices, stemming from human intention or an unintended act, perpetuating the bias(es) for generations. Worse still, because the resulting discrimination is almost always an unintentional emergent property of the algorithm's use rather than a conscious choice by its programmers, it can

be tough to identify the source of the problem or explain it to a court. Machines tend to give the impression that they are neutral, but they are not.

Towards a human-centered perspective of AI

How to develop an ethical and non-biased AI application in an undoubtedly biased and unbalanced society? Can AI be the holy grail by developing more balanced societies that overcome traditional inequality and exclusion? It is too early to say and seems apparent that we will witness many trial-and-error phases before achieving a consensus on what and how AI might be used ethically in our societies.

A fundamental aspect to consider in dealing with issues associated with the development of AI is that AI applications are not another utility that might need regulation once it reaches maturity. AI is a powerful tool that is potentially reshaping how we understand life, interactions, and societal and business environments. All the actors involved must apply new governance systems that allow us to closely monitor and define the role of AI in our future societies in the path of development.

We offer some critical considerations based on the available evidence:

Unbiasing (biased) human beings
Behind the development and implementation of algorithms, there are developers and specific people in power positions. As seen in the data, the developer's professional world is far from being diverse today, explaining some of the thinking logics that foster biases. Increasing the diversity of and access to developer positions in the big companies that dominate the industry would help to increase a more critical perspective of how the algorithms are developed to increase human inclusion rather than the opposite.

The almost monopolistic balance of power in the AI industry may also negatively impact how tech is developing. If we understand algorithmic bias as imposing specific ideas using computers and math as an alibi, we are more able to critically question the institutional logic behind the perpetuation of bias and discrimination practices.

There is a need to increase control and monitoring systems to ensure that human bias does not permeate into how algorithms are created and developed. Recognising the importance of diversity in terms of data and leadership and demanding accountability in certain decisions are essential guiding principles toward achieving a more just development and implementation of AI in the future (Howard and Isbell, 2020).

Data for good instead of data for bias
Algorithms are programmed by human beings, but historical data might carry its own issues related to the accumulation of human biases. In this area, there have been critical advances but more protocols regarding ethical data collection are needed. The existing regulations in some countries might prove insufficient to tackle the issue at present.

Other vital initiatives are developing that might help solve historical dataset biases, such as the one carried out by a researcher at the University of Ontario, who used the MNIST dataset and distilled that database of 60K images down to only 5 to train an AI model (Hao, 2020). Should these procedures be successfully applied to different contexts, they will make AI more accessible to companies that may not be able to afford massive databases and improve data privacy and data collection, as less information from individuals will be required to train relevant models.

Regulation and common ethical frameworks

In a field where the pace of technological development is faster than the ability of regulators to react and create standard guidelines of action and control for the technology, we still need to design common regulation frameworks to ensure an ethical development of AI technologies and to counter the associated risk of concentrating the development of these technologies in just a few organisations.

Educating citizens in the advantages and risks of AI applications

AI development poses diverse and notable challenges with regard to understanding societies, politics, business and even our daily lives as members of society. As AI permeates and increases its role in business processes affecting individuals' choices and possibilities, more education to increase awareness and understanding of these topics is needed.

The technology readiness of citizens will improve AI adoption and have a positive impact on the critical assessment of AI implementation and its effects. A more aware citizen will be less tolerant of manipulation and acceptance of biased or unfair applications of AI tech, such as those related to surveillance that might come into conflict with citizens' liberties and rights.

Conclusion

Making machines more human, or even suppressing human intelligence, has often been treated as one of the ultimate goals of technology advancement. There is a need to understand more about human-machine interactions to safeguard human beings while continuously improving the intelligence of machines. Human-centred technology development implies that the developers and companies using the machines should aim to satisfy the need for innovation while paying attention to their potential impact on society. Humans are flawed. Our society is naturally full of biases that are systematic and institutional, sometimes without us knowing. As we advance the development of society, we should avoid replicating the same issues in the machines we build.

References

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review*, *104*, 671. Available at SSRN: http://www.datascienceassn.org/sites/default/files/Big%20Data%27s%20Disparate%20Impact.pdf

Boulamwini, J (2019). Artificial Intelligence has a problen with gender and racial bias. Here is how to solve it. Available in : https://time.com/5520558/artificial-intelligence-racial-gender-bias/

Cognitive world (2019). How Far Are We From Achieving Artificial General Intelligence? Available in: https://www.forbes.com/sites/cognitiveworld/2019/06/10/how-far-are-we-from-achieving-artificial-general-intelligence/#27f8784b6dc4

European Commission (2018), COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE EUROPEAN COUNCIL, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS Artificial Intelligence for Europe (COM (2018) 237 final). Available in: http://ec.europa.eu/transparency/regdoc/rep/1/2018/EN/COM-2018-237-F1-EN-MAIN-PART-1.PDF.

Garcia, M. (2016). Racist in the machine: The disturbing implications of algorithmic bias. *World Policy Journal*, *33*(4), 111-117.

Hao (2020), A radical new technique lets AI learn with practically no data, *MIT Technology Review*, October. Available in: https://www-technologyreview-com.cdn.ampproject.org/c/s/www.technologyreview.com/2020/10/16/1010566/ai-machine-learning-with-tiny-data/amp/

Howard, A. and Isbell C (2020), Diversity in AI: the invisible men and women, *MIT Sloan Management Review*, Winter Issue. Available in: https://sloanreview.mit.edu/article/diversity-in-ai-the-invisible-men-and-women/

Overgoor, G., Chica, M., Rand, W., & Weishampel, A. (2019). Letting the computers take over: Using AI to solve marketing problems. *California Management Review*, 61(4), 156-185.

Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, 70, 263-286.

Soni, N., Sharma, E. K., Singh, N., & Kapoor, A. (2019). Impact of artificial intelligence on businesses: from research, innovation, market deployment to future shifts in business models. *arXiv preprint arXiv:1905.02092*.

Spice, Byron. 2016. Carnegie Mellon transparency reports make AI decision-making accountable. Carnegie Mellon Computer University School of Computer Science. http://www.cs.cmu.edu/news/carnegie-mellon-transparency-reports-make-ai-decision-making-accountable .

The Verge (2016). Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day, available in: https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist

Vice (2020), An AI Paper Published in a Major Journal Dabbles in Phrenology, available in: https://www.vice.com/en/article/g5pawq/an-ai-paper-published-in-a-major-journal-dabbles-in-phrenology