How to prioritise humans in artificial intelligence design for business

Through the pervasive use of massive amounts of data to automate decisions and processes, artificial intelligence (AI) constitutes one of the most impactful developments for businesses and organisations in general. However, this fast-paced and unstoppable trend raises ethical issues. How can we ensure that AI development is fair, when the algorithms at its core are designed with (often unconscious) racist, sexist, or other biases? **Lorena Blasco-Arcas** *and* **Hsin-Hsuan Meg Lee** *propose a human-centred view for the design of specific frameworks and regulatory systems.*

"Okay, Google, what's the weather today?" "Sorry, I don't understand."

Does the experience—interacting with smart machines that don't respond to orders—sound familiar? This failure may leave people feeling dumbfounded, as if their intelligence were not on the same wavelength as the machines'. While this is not the intention of AI development (to interact selectively), such incidents are likely more frequent for "minorities" in the tech world.

The global artificial intelligence (AI) software market is forecast to boom in the coming years, reaching around 126 billion US dollars by 2025. The success of AI technology is forcing many existing companies to transform their business model and shift to AI. However, along with the advance, there is an increasing worry about the biases in the algorithm development of all these tools.

How AI flaws become apparent

Algorithmic bias is nothing new. However, to date, engineers have focused more on developing AI algorithms to solve complex problems than on monitoring and reporting the potential issues these technological advances bring. We have already seen examples of failures by technology with the rise of discriminatory practices. For instance, in 2016, Microsoft released its self-learning chatbot Tay on Twitter. It was supposed to be an experiment in "conversational understanding." The AI tool could learn language fundamentals, and, over time, it could participate in a conversation by itself. However, the bot ended up developing racist and sexist traits on social media. Another example occurred at MIT when while working on facial recognition, Joy Buolamwini conducted a discriminatory experiment without knowing it. As a dark-skinned woman, she was not recognised by the AI as precisely as her white friend. The results completely missed the point of the experience, and she found out that 99% of white women were identified by the computer, compared to 65% of black women.

Was human intention behind the AI's behaviour? Maybe, maybe not. These examples do not mean that the AI tools were fundamentally flawed or designed to be racist. Nevertheless, their design was biased, and they were not controlled enough before going public. Data biases can lead to discriminatory practices stemming from human intention or an unintended act, perpetuating generations' bias(es). 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 is tough to identify the source of the problem or explain it to a court. Machines tend to give the – false – impression that they are neutral.

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 it 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. 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 by working on three areas:

1. 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, which explains 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 offer a more critical perspective on how algorithms are developed. This would increase human inclusion rather than the opposite. Suppose we understand algorithmic bias as imposing specific ideas using computers and math as an alibi. In that case, we can question the institutional logic behind the perpetuation of bias and discriminatory practices.

There is a need to increase control, monitoring systems, regulation and common ethical frameworks to ensure that human bias does not permeate the creation and development of algorithms. We echo the view of [professors Ayanna Howard and Charles Isbell](#) at Georgia Tech that 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.

2. Data for good instead of data for bias

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. Should these procedures be successfully applied to different contexts, they will make AI more accessible to companies that may not afford massive databases. It will also improve data privacy and data collection, as less information from individuals will be required to train relevant models.

3. Educating citizens in the advantages and risks of AI applications

AI development poses diverse and notable challenges concerning understanding societies, politics, business and even our daily lives as citizens As AI becomes increasingly present in business processes affecting individuals' choices and possibilities, more education is needed to raise awareness and understanding of these topics.

The technology readiness of citizens will improve AI adoption and positively impact 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 conflict with civil liberties and rights.

Making machines more human, or even suppressing human intelligence, has often been treated as one of the ultimate goals of technological advancement. Human-centred technology development implies that the developers and companies using the machines should not only aim for innovation but also pay attention to their potential impact on society. Humans are flawed, meaning our society is naturally full of biases that are systematic and institutional, and we are not always aware of them. But we should avoid replicating the same issues in the machines we build.

<div align="center">♣♣♣</div>

Notes:

- *This blog post is based on [AI, Friend or Foe? A Proposal for a Human-Centered Approach to AI Business Implications](#), part of ESCP Business School's ["Better Business: Creating Sustainable Value" series](#).*
- *The post represents the views of the author(s), not the position of LSE Business Review or the London School of Economics.*
- *Featured [image](#) by [Shopify Partners](#), under a [Burst](#) licence*
- *When you leave a comment, you're agreeing to our [Comment Policy](#).*