‘AI dangerously biases our decisions’

Debates and Ethics of Big Data and Data Science

Jose Antonio Lorencio Abril, Rishika Anil Gupta

3rd March 2023

Abstract: "We already have many well known and documented examples of how AI and ML models introduce biases. It is also true that human perceptions are biased by media as well as cultural and social cliches but the bias introduced by data-driven systems is more dangerous from a social point of view."

General Context:

Artificial Intelligence is an integral part of our lives. It is constantly evolving from voice automation devices in the household to vehicles without drivers. Computers, although being dumb devices, have the capabilities like humans to make decisions. In the near future, we might have delivery systems or agents that deliver couriers via drones and it is Artificial Intelligence (AI) that can turn these ideas into reality. However, people often debate about the genuineness and trustworthiness of foolproof-ness (as questioned by the jury) of these AI systems in decision making.

Opposition Criteria:

AI systems are learning from real-world data, and the objective is to identify patterns in it, so that we can better understand the nature and behavior of the sources of data. If these sources have biases, which is common, because data is usually produced by human actions and relationships, these biases can be introduced into the AI system. Our claim is, that even though these biases can arise, they are not necessarily dangerous, and can be used to actually improve the current state of things, and even to revert those situations that we consider unjust.

Points of Arguments:

- Quoting Pieter den Hamer the Sr. Director and Analyst at Gartner who states - *“With continually more dynamics and complexity in modern-day business, our capabilities must improve to make the best possible decision in the shortest possible time, in a scalable, risk-conscious, consistent, adaptive and personalized fashion”* [1]. We believe that decision-making is a key component in each of our lives, like as of tonight we need to pick a specialization institution to continue our Semester 3 and 4. There are many factors - curriculum, housing, research, environment, etc. that might cater to this decision and complexity of each factor varies from person to person. In the context of any IT organization, they take technical decisions like - deciding which software to use, what databases should be adopted, etc. In all such contexts, AI will make more effective decisions with or without bias, as it clearly knows when and why to complement the best of human decision making with the power of data and analytics and artificial intelligence [1].

- In the same line of reasoning and considering the point made by the **proposition team**, that humans can’t handle huge data, we believe that it is when AI comes to rescue, as these models can be used to analyze large amounts of data and provide insights that may not be apparent to humans, leading to better-informed decision-making, justifying that AI is a great tool that aids to make decisions by adding a value to the decision process that would simply not be possible without it. A remarkable example of this is the usage of AI in medicine, which is helping doctors to identify illnesses in patients that the doctors would classify as healthy. This is because AI can leverage the huge amount of data that can be measured from a person, making connections between variables that are too intricate for us to notice or comprehend easily [2], [4]. The **proposition team** also stated that bias can decide if someone dies or lives because of its bias towards some social group, exemplifying this by how in healthcare a system could provoke a bad outcome because of this bias. Nevertheless, we have seen how this kind of system is helping doctors to better identify illnesses, thus saving the lives of millions of people. They did not provide a specific example, but as of our knowledge, this is anecdotal and the benefits of using the systems outweigh the cost. Also, it is

always the doctor who makes the final decision, so if a doctor does not detect anything with the aid of an AI, it is unlikely that the same doctor would have detected it without this aid.

- Not only this, but also AI can leverage the bias identified in data to actually revert unjust situations. Even if an organization has a short-term lucrative incentive to use biased patterns detected, in the following example we can see how this behavior is socially punished and not allowed at all. The **proposition team** presented the Cambridge Analitica example, in which this is perfectly exemplified; it was proven that they used data to gain profits, which led to the society reacting accordingly and ultimately to the closure of the company. This means that companies must incorporate good practices in their decision making processes, and it is the values of transparency, responsibility & respect that are regarded as desirable by consumers.

- Furthermore, humans are susceptible to emotions, conditions as well as circumstances, which have a profound effect on decision-making abilities. Like on a good day, we are bound to take rational decisions while on a bad day or unseen circumstances, we are bound to take irrational decisions. On the contrary, AI models are objective in nature (unlike humans - subjective) and do not introduce bias in systems, thereby having an upper hand than humans on improving the decision-making process.

- Each person has their own narrative (the way a sentence is said or meant, like for example - stressing on pauses or commas in a sentence can bring out a different meaning in the sentence entirely). AI models are programmed by skilled people considering all the possible edge cases and at the same time incorporating all narratives, which is why there is no chance of introducing biases in decisions. Even if the programmer introduces bias (catering to their own perspective or narration that might harm the quality of decisions), before deploying AI models on production, they are thoroughly reviewed by an experienced group of managers/committee members, who are ethically as well as lawfully unbiased. In either case, AI models are adding value to the process of decision making. The **opposition team** remarked that AIs are as unbiased as the humans that created them, but we take the argument further, stating that in fact they are less biased than us, because they have been designed through an engineering process, in which the engineers perform thorough analysis and tests, thinking in a long-term perspective, which tends to be more objective than our day-to-day decisions.

- Continuing the points made above, it is in our hands, as we can program AI to recognize and eliminate bias in data and algorithms, which shall lead to more accurate and fair decision-making processes, thereby reducing discrimination against certain groups. For instance, a study found that an AI system could reduce racial bias in criminal sentencing by providing judges with more accurate information about defendants [3]. Not limited to this, but efforts are being made to incorporate the awareness of bias into the training processes of AI systems [6].

- Talking about biases in general, as the **opposition team** discussed, they come from humans and are our mirrors and have been molded into our lives since childhood by our teachers, parents, friends, etc. AI models are trained on existing data, meaning the model understands the data better. The bias exists by nature (not on purpose), which is why models are trained correctly and better, and can actually help us to take consciousness of these biases.

- The **proposition team** asserted that AI is everywhere nowadays, and that it is affecting our daily lives and decisions, without us even noticing. This is true to some extent, but it is also true that it is an individual responsibility to use the tools that are available to mitigate this impact, and to be aware of the implications that the use of technology has in our lives. A simile is encountered in the context of news. Newspapers have their own biases, and it is the responsibility of the reader to interpret the text, in order to minimize the effect of the newspapers' biases or agenda in his/her own life.

- On the lines of the previous point, we agree to some extent that AI algorithms can introduce biases in decision-making, but they help us make better-informed decisions and improve outcomes. Other than the examples of healthcare mentioned in the above points, in the finance industry, AI is used to identify fraud

and reduce risk, while in manufacturing to optimize production and improve quality control. In each of these cases, the benefits of AI outweigh the potential risks and biases.

Final Considerations:

We, as aspiring data scientists, are going to be developing and using AI systems throughout our lives. It is our responsibility to be aware of all the potential risks of these technologies, and leverage them to make good and to minimize the harm caused to others. This does not mean that we should defend that AI systems are dangerous, but rather that they can be misused if their limitations are not understood or are disregarded. It is our duty to design our systems to be robust and as unbiased as possible, as well as to communicate how they can be used, and how they should not be used.

In conclusion, AI systems can be used to better understand society and to correct unjust situations if the developers and the decision makers are aware of the limitations of AI. Since we have seen the huge benefits and incentives for the responsible use of AI, we can infer that we are on the good path and that these systems are going to help us construct a more fair society.

References:

[1] https://www.gartner.com/smarterwithgartner/would-you-let-artificial-intelligence-make-your-pay-decisions, visited on 3rd March 2023, 14h50m.

[2] Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. Nature medicine, 25(1), 44-56. url: https://www.nature.com/articles/s41591-018-0300-7

[3] Ensign, D., Friedler, S. A., Neville, S., Scheidegger, C., & Venkatasubramanian, S. (2018). Runaway feedback loops in predictive policing. ACM Conference on Fairness, Accountability and Transparency (FAT), 289-298.

[4] Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. New England Journal of Medicine, 380(14), 1347-1358. url: https://www.nejm.org/doi/full/10.1056/nejmra1814259

[5] Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., ... & Dean, J. (2019). A guide to deep learning in healthcare. Nature Medicine, 25(1), 24-29.

[6] Sarah Bird, Krishnaram Kenthapadi, Emre Kiciman, and Margaret Mitchell. 2019. Fairness-Aware Machine Learning: Practical Challenges and Lessons Learned. In Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining (WSDM '19). Association for Computing Machinery, New York, NY, USA, 834–835. url: https://doi.org/10.1145/3289600.3291383