

Article 2 - Addressing Bias in AI: Challenges and Solutions

Introduction

AI (Artificial Intelligence) has become a part of our daily lives—be it in hiring, healthcare, banking, or even social media. But did you know that sometimes AI can be unfair? It can make decisions that unintentionally discriminate against people. This is what we call **bias in AI**, and it's a big problem.

Let's understand this better with some real-world examples that I discussed in [Article-1](#). I'll also share what can be done to fix these issues.

1. Hiring AI Prefers Men Over Women

In 2020, a recruitment AI tool started picking more men than women for jobs. This happened because the AI was trained using past hiring data, which already had a bias.

In 2021, a credit card company used AI to decide credit limits. Women were getting lower limits even though their financial profiles were the same as men's.

Solution: The company has to re-train the AI with the properly selected sample data, Company should consider re-training of AI at least once in a year with new data.

2. Healthcare AI Ignoring Black Patients

A healthcare tool in the US gave more priority to white patients than Black patients for treatments. It assumed spending more money on health meant the patient needed care, which wasn't true for all communities.

Solution: System should be focused on the important parameters such seriousness of the illness, type of illness, not spending patterns. Use **SHAP** technique for to define the parameters for the training. This has to be documented and reviewed by industrial SME, in this scenario doctors from various sector.

3. Moderating LGBTQ+ Content on Social Media

Some social media platforms flagged LGBTQ+ posts as inappropriate due to biased keywords.

Solution: AI team have to work with LGBTQ+ groups for better understand the content and fine-tuned the AI systems.

4. Voice Assistants Struggling with Accents

Voice assistants like Siri and Alexa didn't understand South Indian or African accents well because the training data didn't include them. I personally faced this issue as it will never detect my voice if i am trying to call my wife using voice!!!!...

Solution: The data collection should be diverse or possibly they can outsource the local data collection or training of AI to the local company which can solve the problem also creates new possible job opportunity.

What Can We Learn from These Examples?

From all these examples, we see one thing clearly: **AI learns from the data we give it.** If the data has biases, the AI will also have biases. But this can be fixed! Here are some simple steps:

1. Use **diverse and inclusive data** to train AI.
2. Conduct regular **fairness audits** to check for bias.
3. Always keep **human oversight** in decision-making.
4. Follow strict **ethical rules** when building AI systems.
5. Involve the community and experts in the industry to understand the real-world impact.

Conclusion

AI is like a mirror—it reflects the data and decisions we feed into it. If we want AI to treat everyone fairly, we must take responsibility for its fairness. By learning from these challenges, future we can create better systems that respect and serve everyone equally.

Bias in AI is not an unsolvable problem—it just needs our attention, care, and effort. Let's work together to build AI systems that are fair and inclusive for all!

Revision #2

Created 17 December 2024 18:40:48 by Admin

Updated 17 December 2024 19:07:03 by Admin