The 17th Annual Social Informatics Research Symposium and the 3rd Annual Information Ethics and Policy Workshop: Sociotechnical Perspectives on Equity, Inclusion, and Justice (SIG-SI and SIG-IEP)

Game based training as a model for skill enhancement in bias mitigation efforts

Sundar Narayanan
Independent Researcher, India
Sundar.narayanan@aitechethics.com

ABSTRACT

Organizations which are developing or deploying machine learning models, have an inherent need to enhance their bias mitigation systems to minimize harms that their models may contribute to. Currently, in organizations, bias mitigation is undertaken by addressing data quality, streamlining process, and structuring appropriate performance metrics for models. However, these approaches do not contribute to the skill enhancement specifically with reference to bias perception, understanding and treatment for people working in developing or deploying machine learning models. To that end, the paper proposes, game based intuitive method to complement and enhance the skills of people.

KEYWORDS

Bias Mitigation, Skill enhancement, Machine Learning

INTRODUCTION

Algorithmic Bias is a widely debated subject currently specifically on the harms contributed by it to people. The harms could include psychological, physical, legal, social, and economic harms (University of Virginia). Algorithmic Bias essentially is the scenario wherein the algorithm systematically prejudices individuals/ groups in its results. This is caused by the data (bias in data collection and annotation), the algorithm (bias contributed by model development), and/ or the outcome impact (bias contributed by outcomes and where such model is used).

These biases exist across the machine learning stages including biases in (a) framing the problem, (b) collecting the data, (c) preparing, or processing the data, (d) model and parameter choices and (e) optimizing the model. (Karen, 2019). For instance, data normalization or outlier removal in the training set can also; in some cases; lead to algorithmic bias. Further there could be instances of non-response by select group in training set can contribute to bias or in some instances de-duplication can also lead to algorithmic bias.

This paper explains the need for skill enhancement, expresses the context of use of games for skill enhancement and exhibits the need for skill enhancement in bias mitigation context. It further draws out instances of games from customized chess games to visual observation games.

84th Annual Meeting of the Association for Information Science & Technology | Oct. 29 – Nov. 3, 2021 | Salt Lake City, UT. Author(s) retain copyright, but ASIS&T receives an exclusive publication license.

SKILL ENHANCEMENT FOR BIAS MITIGATION

Addressing bias for the organizations is a layered set of actions which involves (1) addressing data quality issues, (2) streamlining model development and decision process, (3) structuring appropriate metrics and, (4) having a more inclusive and diverse feedback loop. (The Center for Equity Gender and Leadership, 2020). These are currently implemented or explored using workflows, machine learning pipeline or ML ops tools, checklists or guidelines and review/ oversight mechanisms. (Shea Brown, 2021). Model cards (Mitchell-et-al, 2018) and Datasheets for Datasets (Gebru-et-al, 2020) are becoming guidelines for people in development and deployment of AI systems to use in the process of bias identification and mitigation, amongst other uses of such guidelines. Further, multi-stakeholder feedback is emerging to be another key contributor to the process; however, the diversity is difficult to achieve for many organizations currently. (Nicol Turner Lee, 2019).

However, one of the factors that is rarely spoken about is the need to enhance skills relating to bias perception, bias understanding and bias treatment for people who are in the development and deployment of algorithms. An important question is whether the adoption of layered set of actions referred ensure bias mitigation? The answer will exhibit subjectiveness tied to the need for skill enhancement for people in development and deployment of AI systems. Bias mitigation is an essential skill in today's machine learning development process that complements with existing organizational processes and technology.

Bias perception, understanding and treatment are impacted by skewed interpretation or perception of the data, fractional time in data processing and analysis in the model development process and prioritizing performance metrics to align with stakeholder expectations. These are further accentuated by cognitive biases people developing or deploying machine learning systems. Some of the root causes of such cognitive biases are emotional and moral motivations, social influence, use of information processing shortcuts, not being aware of vanishing options and limitations in ability to processing information. (Desjardins, 2017). Skill enhancement therefore shall focus on intuitive engagement with professionals that helps them overcome cognitive bias and contribute to mitigating bias.

USING GAME BASED EFFORTS FOR SKILL ENHANCEMENT

Games are very effective in the learning process. This is because they create an environment that makes the learner to get engaged and influence cognitive decision-making process through experiences and metaphors. It helps in following ways:

- Embedding constrains and preparing for in real world scenario.
- Enhance problem solving ability.
- Exhibiting indicative consequences of their actions.
- Support in assimilating the overall big picture and improve creativity
- Trigger emotions and experiences through the game.

While games are known to be an established approach for learning in many fields, this paper proposes an approach for using games for Bias mitigation. The author argues that this can be considered as a effective tool for enhancing bias awareness and attempting more informed or involved approach towards bias mitigation.

Illustrative Games

Sl no	Game	What it does?	How does it help?
1	Chess Variant	Variants of chess where the powers of queen and the king are swapped or the score for the pawns, thereby discouraging the players to lose pawn.	Exposes the challenge of discriminated society and the underlying impact on people.
2	Bias Spotting	Multiple images are placed, and the player is asked to spot the bias out of them.	Exposes the bias embedded in the society and creates awareness of various types of societal bias.
3	Fact checks games	Players are provided certain content representative of social media news, for them to consume. The players must spot which of those news contents are fake.	Creates awareness to differentiate inconsistencies caused by fake news or misinformation (also exposes

			awareness to multi-stakeholder feedback as an attack vector).
4	Graph Gaps	Players are provided with several graphs in different ways, and they must spot the approximation errors and misrepresented graphs.	Enables perception of inconsistent graphs and how one views a graph can make a difference in spotting a bias.
5	Role Play Games	Players are provided with a role play game wherein they must collect balls (different colour, shapes, size and types) from different locations to finish every level. They also earn by collecting coins. At advanced levels they need to exchange their coins for balls from characters. The more diverse balls they can collect, they get more points.	Enables better perception of diversity in dataset.
6	Choice	Players are asked to make certain choices as they enter the game and asked to revisit the choices as they progress. They can drop a choice and pick a new choice to navigate through the game. This is based on quandary model.	Helps understand the power of choice and opportunity to revisit choices for better outcomes.
7	Moral dilemma	Players are provided with moral dilemma wherein they need to provide what option will they choose in the dilemma	Provides opportunity to experience the dilemma and inherent implications of the choice.

.

CONCLUSION

Skill enhancement is critical as it elevates the base line of the problem progressively over period, resulting in continuous improvement. Games exhibits constrained environment and enables cognition of the underlying challenges. We propose AWE games that help in creating bias Awareness, influencers Willingness to address them and enables bringing Equitableness in the models. This approach does not provide a decentralized approach of regional teams to adopt approaches that may help in addressing bias in regional environments.

Further, using games as a model complement with the internal processes and monitoring metrics that organization have, but also enables collective pursuit of the people developing and deploying machine learning models to minimize harms caused by bias in models.

REFERENCES

1. Desjardins, J. (2017). Every Single Cognitive Bias in One Infographic. Retrieved from https://www.visualcapitalist.com/every-single-cognitive-bias/

2. Gebru-et-al. (2020). Datasheets for Dataset. Retrieved from https://arxiv.org/pdf/1803.09010.pdf

3. Karen, H. (2019). This is how AI bias really happens—and why it's so hard to fix. MIT Technology Review. Retrieved from https://www.technologyreview.com/2019/02/04/137602/this-is-how-ai-bias-really-happensand-why-its-so-hard-to-fix/

4. Kim, S. (2021). AI and Robots Are a Minefield of Cognitive Biases. IEEE Spectrum. Retrieved from https://spectrum.ieee.org/humans-cognitive-biases-facing-ai

5. Mitchell-et-al, M. (2018). Model cards for Model reporting. Retrieved from https://arxiv.org/abs/1810.03993

6. Nicol Turner Lee, P. R. (2019). Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms. Brookings . Retrieved from https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/
7. Shea Brown, R. C. (2021). Bias Mitigation in Data Sets. ForHumanity. Retrieved from https://static1.squarespace.com/static/5ff3865d3fe4fe33db92ffdc/t/60df26501d27325c69703a30/1625237072384/biasInDatasets+%282%29.pdf
8. The Center for Equity Gender and Leadership. (2020). Mitigating Bias in AI, An Equity Fluent Leadership Playbook. Haas School of Business (University of California, Berkeley). Retrieved from https://haas.berkeley.edu/wp-content/uploads/UCB_Playbook_R10_V2_spreads2.pdf
9. University of Virginia. (n.d.). Types of Harms. Retrieved from https://research.virginia.edu/types-harm