Published in Towards Data Science

Jillian Chambers

Dec 22, 2021 · 5 min read · Listen

Can Algorithms be Racist?

A brief exploration of juristic and social dilemmas within language models

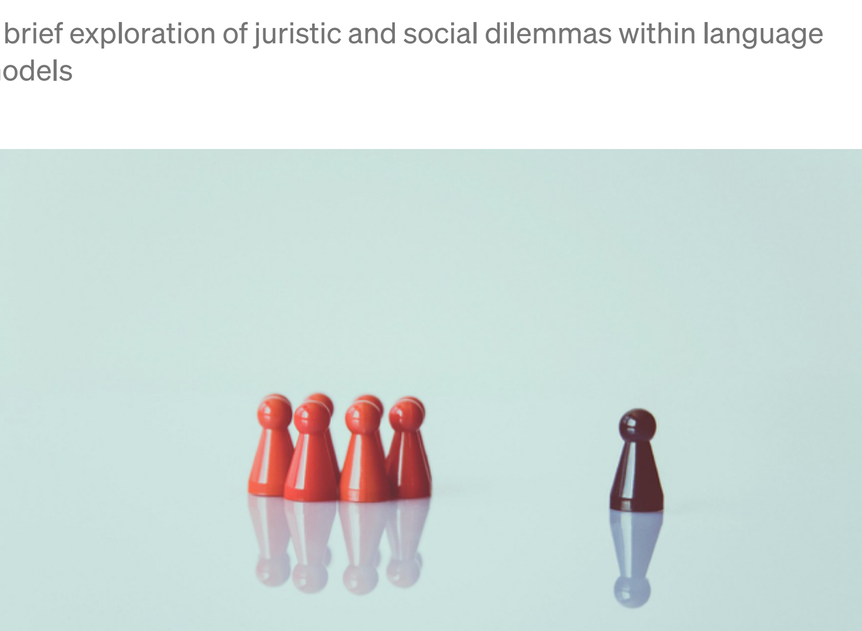

Photo by Markus Spiske on Unsplash

As artificial intelligence (A.I.) continues to rapidly integrate within everyday life, there are a few ethical dilemmas that have arisen synchronously and their impact on use cases have become the subject of much debate (Kilbertus et al., 2017; Hardt et al., 2016; Pazzanese, 2020). One such predicament that this paper hinges on has to do with inclusivity and marginalization (Bender et al., 2021). How are notions of participation affected by training data that reinforce hegemonic power in the formation of algorithmic models? Accordingly, this article will seek to spotlight ethical challenges within A.I. via a grounded interpretivist viewpoint gained by qualitatively investigating the literature in order to discuss bias amplifications. As outlined by Bender et al., (2021), there are several juristic and social dilemmas regarding the growth and utilization of language models. This includes the reinforcement of hegemonic power over marginalized populations as a result of training data characteristics that reify power imbalances.

Affluent countries benefit more from language models while poorer countries suffer the negative environmental ramifications of training data that barely benefits them (Pratik et al., 2020); they are also largely excluded from or underrepresented in technological advancements. In addressing these issues, Bender et al., (2021) propose the careful planning and curation of datasets with a view to adopting a value sensitive design approach in order to reduce the negative impacts caused by language model inaccuracies, and the perpetuation of stereotypes that further subjugate certain groups.

Additionally, it was recommended that benchmarks be created when constructing models that would critically scrutinize data-gathering and usage via premortems to increase inclusivity, ensure fairness and encourage accuracy. Notions of participation are also negatively affected by training data that is contextually and institutionally grounded in bias within deep learning. For example, the algorithm used in Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) in the U.S., has an inherent bias toward African-Americans, as it gives this group a higher risk rating than Caucasians, even when both have the exact offender profile (Mehrabi et al., 2019). Obermeyer et al. (2019) references a popular algorithm widely used within US hospitals that lessens the likelihood of Black people accessing medical programs for the same complex health situations as other groups.

This reinforcement of hegemonic power in the formation of deep learning algorithmic models spotlight bias amplifications within artificial intelligence, and the need for rigorous checks to identify, uproot and discard training data that are discriminatory, whether during the data gathering, pre-processing, in-processing, or post-processing stages (Mehrabi et al., 2019). Computer vision is another area of A.I. where hegemonic structures exist. Buolamwini et al., (2018) highlight issues with gender and race in algorithms that analyze faces: males with lighter skin-tones had less instances of assignment errors than females with darker skin-tones. Once again, how algorithmic models arrange the visual world in a way that forms a standard native intelligence depends on the datasets it receives which often is not inclusive.

For example, a quick Google Images search for coronavirus will show a plethora of images of colorful molecules, while a search for Ebola will output pictures of various African peoples in debilitating conditions. Although coronavirus has had a greater negative impact globally, the images online do not reflect as badly on one group of people as opposed to another. One can then pose the question: who determines what should be seen and how? Who gets to decide how the actual and artificial worlds are negotiated? If computer vision therefore has the power to shape perception which consequently patterns behavior and thought, then, as Pereira (2021) suggests, there is a need to radically break away from these discriminatory constructs. But how can this be done? What motivations are there for researchers to be minded toward ethical challenges within the aforementioned areas of A.I.?

Firstly, financial subsidies should be granted to researchers whose work(s) actively demonstrate equilibrium of power dynamics- that is to suggest the elimination of ascendancy of one demography over another overtly or covertly- within the training data or deployed model. Secondly, and indeed conversely, economic support or public funding should be removed, and algorithmic models created invalidated as an aversive stimulus for AI researchers who do not take heed to ethical dilemmas in their investigations. Thirdly, a peer-reviewed framework of standardization and taxonomy of best practice metrics should be established from high-quality research that actively and consistently incorporates the use of ethics throughout the research process.

A.I. is used successfully across major industries and for that reason, the aggregate exposure to various challenges within this domain is all-encompassing. Considering this, if datasets that train algorithms are intricately confined to certain hegemonic principles, then the datafication process will ultimately augment this, further challenging notions of participation. Ethicality is therefore of critical importance within A.I. and researchers have a duty toward ensuring that concepts of wrong and right, fairness, and balance are defended and advanced systematically.

Bibliography

Bender, E., Gebru, T., McMillan-Major, A & Shmitchell, S. (2021) On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?. C*onference on Fairness, Accountability, and Transparency.* New York, 3–10 March, 1–14.

Brownlee, J (2019). Multi-Label Classification with Deep Learning. Available Online: https://machinelearningmastery.com/impact-of-dataset-size-on-deep-learning-model-skill-and -performance-estimates/ [Accessed 12/12/2021].

Brownlee, J (2021). How to Choose an Activation Function for Deep Learning Available Online: https://machinelearningmastery.com/impact-of-dataset-size-on-deep-learning-model-skill-and -performance-estimates/ [Accessed 12/12/2021].

Buolamwini, J. & Gebru, T. (2018) Proceedings of Machine Learning Research 81:1–15.

2018 Conference on Fairness, Accountability, and Transparency Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. New York, 23–24 February, 1–15.

Doshi, S. (2019) Various Optimization Algorithms For Training Neural Network. The right optimization algorithm can reduce training time exponentially. Available Online: https://towardsdatascience.com/optimizers-for-training-neural-network-59450d71caf6 [Accessed on: 12/12/2021].

Dabbura, I. (2017) Gradient Descent Algorithm and Its Variants. Available Online:https://towardsdatascience.com/gradient-descent-algorithm-and-its-variants-10f65280 6a3 [Accessed at: 14/12/2021].

Hardt, M., Price, E., Srebro, N., Equality of opportunity in supervised learning. *30th Conference on Neural Information Processing Systems (NIPS),* Barcelona, 5–11 December, 3315–3323.

Images (no date). Available at: https://pixabay.com/ (Accessed 13 Dec 2021).

Kilbertus, N., Rojas-Carulla, M., Parascandolo, G., Hardt, M., Janzing, D & Sch ¨ olkopf, B. (2017) Avoiding discrimination through causal reasoning. *31st Conference on Neural Information Processing Systems (NIPS)*, California, 8/6/2017, 656–666.

Nguyen, H & Evangelos, T. (2021) The Impact of Overfitting and Overgeneralization on the Classification Accuracy in Data Mining. PhD Dissertation Louisiana State University.

Obermeyer, Z., Powers, B., Vogeli, C & Mullainathan, S. (2019) Dissecting racial bias in an algorithm used to manage the health of populations. *Science, 366*(6464), 447–453.

Pazzanese, C. (2020) Ethical concerns mount as AI takes bigger decision-making role in more industries. Available online: https://news.harvard.edu/gazette/story/2020/10/ethical-concerns-mount-as-ai-takes-bigger-de cision-making-role/ [Accessed 24/11/2021].

Pereira, G. (2021) Towards refusing as a critical technical practice: struggling with hegemonic computer vision. *APRJA* 10(1), 2245–7755.

Pratik, J., Sebastin, S., Amar, B., Kalika, B & Monojit C. (2020) The State and Fate of Linguistic Diversity and Inclusion in the NLP World. *58th Annual Meeting of the Association for Computational Linguistics*. Online, 6282–6293.

Rohith, G. (2018) Improving the Performance of a Neural Network. Available Online: https://www.kdnuggets.com/2018/05/improving-performance-neural-network.html [Accessed 13/12/2021].