Sagarika Suresh Thimmanayakanapalya -50320276

UB School of Management

**Writing: “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?”**

The section on the unfathomable training data speaks of how the size of the data doesn’t particularly guarantee diversity. This particularly interested me because, although the size of the training data can be huge while training language models, if the data is not diverse enough, it cannot handle marginalized populations. However, I do feel diversification of data depends on the type of context and problem that it is used to address. For instance, when a language model has to be trained according to a given website in some research areas, such as a bot that has to replicate a given website then curation according to the context plays a bigger role than curation in general. The authors also suggest that the training data must be filtered to avoid the risk of perpetuating dominant viewpoints. I believe that this point must be further elucidated, as filtering or curation of datasets could also lead to other types of bias. For instance, questions such as who is filtering the data? What measures are they taking to filter it? and why are they taking it? must be strongly addressed. I strongly feel that the dataset must be meticulously curated such that it ensures that other biases do not crop up during curation.

Another important issue that I felt with collecting large datasets for training language models is that how do researchers make sure that the language model does not go against the community guidelines [1]. For instance, in healthcare, patients have compliance issues, however, a lot of data for language models is collected from medical transcriptions. When a lot of data is being collected through web crawlers and web scrapers to train language models, researchers must look into ways in which the data scraped to train these language models are following compliance issues and not invading an individual’s privacy. Finally, I would like also to add that the paper did not talk about slang and context-specific to different geographical locations. For instance, the paper speaks of how large volumes of data cannot be diverse but it does not speak of the varying ethnographic conditions present in different parts of the world, that would also affect the day-to-day lives of marginalized groups [2]. These could be cases where certain manners in which Alexa speaks to certain ethnicity could be considered either rude or polite based on where the pre-trained was collected. So, I would personally lay more stress on the particular context while training language models.

**References:**

**[1]** Pan, X., Zhang, M., Ji, S., & Yang, M. (2020, May). Privacy risks of general-purpose language models. In *2020 IEEE Symposium on Security and Privacy (SP)* (pp. 1314-1331). IEEE.

**[2]** Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*.