Homework - 1

After reading this paper, I find the 'Size Doesn't Guarantee Diversity' section most interesting because I always believe that all NLP models consist of all kinds of information. But in reality, only dominant views are considered in the model. And that creates a difference of opinions between communities.

In this paper, I agree with all the mentioned points, but I am giving two reasons to be more specific. The first one is training any large neural networks can be very costly. For instance, GPT-3 comprises 96 layers and 175 billion parameters. To train such model 3.14E23 FLOPS[1] requires approximately 355 years, and the cost is around $4.6M for a single training run. Because of that, the overall cost of the product is increased. And now GPT-3 is available publicly; not all people in the world can afford this product. Plus, GPT-3 is nothing but a big transformer with massive computational power, an enormous amount of electricity, and several resources like servers, GPU, etc. Such machine learning technologies are incredibly energy-intensive, and according to the given paper, it emits 248t of CO2, which has harmful effects on the environment. So, training large models is costly not only financially but also economically.

The second reason is that even though this model has 175 billion parameters still, it is biased towards gender, race, and religion; also, it sometimes spreads fake news all over social media. For example, in the paper "Gender and Representation Bias in GPT-3 Generated Stories[2]," it is mentioned that most of the generated data are more likely to have masculine characters than feminine characters. Also, the present feminine characters are mostly related to the terms like life, family, and appearance, whereas masculine terms are related to politics, war, and machines. And this shows gender biases in the model.

Moreover, the previous version of GPT-3, which is GPT-2, contains undesirable language such as slurs, offensive and threatening speech primarily related to minority groups (e.g., words such as "gay"," Muslim"). The paper "Detoxifying Language Model Risks Marginalizing Minority Voices[3]" evaluates that compared to text containing White-Aligned English, the detoxification substantially increases the large model's perplexity on text with African American English. It shows the race and religious biases in the model.

From my point of view, GPT-3 is adversely affecting social media. In day-to-day life, NLP-based applications like Facebook, Twitter, and Instagram are used to do illegal stuff by creating fake accounts. For example, GPT-3 contains GAN[4] functionality, which some people can use to generate new, synthetic data that can pass for real data. In July 2021, an independent data scientist researcher who goes by "Conspirador Norteño" and their partner "Dr. ZQ" discovered a bot network on Twitter of more than 3000 fake accounts using GAN-generated images. These fake accounts are used for Pornography and Online sports betting as it is a clear violation of Twitter's terms of service.

In concluding, I believe that the authors have included every dangerous aspect of the large model by considering marginalized people's points of view.

Citations

1. (Li, 2020) https://lambdalabs.com/blog/demystifying-gpt-3/
2. (Li Lucy, 2021) https://aclanthology.org/2021.nuse-1.5.pdf
3. (Albert Xu, 2021) https://arxiv.org/pdf/2104.06390.pdf
4. (Gault, 2021) https://www.vice.com/en/article/z3v579/twitter-accounts-with-ai-generated-cat-avatars-at-center-of-turkish-porn-bot-ring