Stochastic Parrots

Isaac Boaz

May 29, 2024

Abstract

This paper will introduce the concepts and ideas that the paper (the paper) 'On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?' by Emily M. Bender and Timnit Gebru.

For its introduction, the paper poses its core question: **How big is too big?** This isn't a question of the size of the model, but rather challenging the downsides (both direct and indirect) of language models (LMs), and how these downsides may disproportionately scale with the size of the model.

The paper introduces the following key points:

- The environmental impact of LMs
- The disproportionate benefit LMs offer to marginalized communities
- The source of the data used to train LMs
- The documentation of training data
- The understanding of the limitations of LMs

Background

The paper briefly covers the history of LMs, pointing out that they were 'proposed by Shannon in 1949', with the earliest implemented ones dating back to the 1980s. The paper

points out how historic models didn't necessarily perform better when increasing the number of model parameters, and we only saw this trend occurs with transformer models (in contrast to n-gram ones).

Cost

The paper goes on to discuss the environmental impact of LMs, particularly with the increased size of transformer models. Specifically, the paper reports that training a large LM model emitted 284t of CO_2.

The paper also mentions how while some of the energy used to train these models may be renewable, it points out

> renewable energy sources are still costly to the environment, and data centers with increasing computation requirements take away from other potential uses of green energy.

The paper then points out that though this cost generally impacts the global population, the benefit is primarily towards the privileged few. In short, the paper proposes 'These models are being developed at a time when unprecedented environmental changes are being witnessed around the world'.

Training Data

Revisiting the source of the data used to train LMs, the paper fleshes out the issues relating to "stereotypical and derogatory associations along gender, race, ethnicity, and disability status". The issue relies on who has access to the internet, and who takes the time to contribute to the data. As a result, the paper explains, that white supremacy, misogyny, ageism, etc. are 'overrepresented' in the training data.

In short, the paper argues that the sourcing, maintenance, and updating of the data is misrepresentative of the world's population.

Parroting

The idea of 'parroting' is introduced as the idea that LMs "amplify biases and other issues in the training data". The paper argues that this 'parroting' is primarily enabled by human error and bias, such as the metrics used to measure the effectiveness of a model, or the data used to train the model.

One core issue, the paper argues, is that LMs lack an understanding of who they are talking to, and vice versa. Human-to-human communication is more obviously based on shared understanding, and even an author writing a book has a sense of who their audience is. LMs, however, lack this understanding. As a result, the paper argues, LMs are not grounded in communicative intent. Thus, the illusion of comprehension is due to the reader's understanding of language.

Risks

Revisiting a core issue that the paper raises, the paper reiterates the idea of garbage in, garbage out. The paper reiterates that the source of the "overrepresented" privileged people's data implies that their views (i.e. racism, misogyny, ableism, etc.) are similarly overrepresented.

Other malicious activities like phishing, spam, and disinformation are also a risk, the paper argues. In addition to malicious activities, other tasks commonly assigned such as translation are "inaccurate yet both fluent and [...] coherent in its own right".

The paper also argues that ignorant or uninformed use of LMs is a major risk, bringing up an example of a person who was arrested due to a faulty MT translation.

Lastly, another risk that LMs provide is the potential for abuse of the underlying training

data. Two examples that the paper provides are discovering personally identifiable information (PII), and querying for illegal or malicious content (such as tax avoidance).

Summary

In summary, the paper argues that the fact that LMs can pass off as 'comprehensive' without understanding the underlying data, or the lack thereof, is a major risk. The paper recaps the risks, costs, and benefits of LMs, and argues that major work is needed to address these issues.