Large Language Models' role in the spread of Fake News

Niteesh Saravanan, Luke Sansonetti, Jai Pandia, Robin Gould

Generation of Fake News

- Large Language Models (LLMs) can be trained to generate articles designed to mimic reliable pieces of news
- A.I. algorithms can be trained to determine if an article is written by a large language model, however we are yet to develop a reliable way to detect whether a pieces of news is written by a human or A.I.
- While accurate A.I. detection software is actively being researched,, it has proven to be much more difficult, and much less profitable than creating LLMs.
- Large language models continue to quickly develop and become more advanced every day, while detection software is still very rudimentary and unreliable.

Why is this important

- LLM can generate fake news quickly and easy and enable spread at scale
- Fake news is mostly spread online where people are more likely to believe it without verifying its legitimacy
- Fake News created by LLMs can be incredibly difficult to detect due to variations in the prompt or model used. We currently have a difficult time already detecting LLM generated content (the word delve is being used as a metric)
- Detectors cannot reliably fact check news, so they are overly biased against LLM generated content which may be real news (or potentially ESL)
- LLMs can enable bad actors to easily create propaganda campaigns that have a massive reach in the places people are most likely to fall for them

Who does it affect

- The problem as mentioned before is that LLMs can generate text so well written that it's hard to differentiate between it and a professional's.
- Additionally, it can replicate the voice of nearly anyone given sufficient data.
- As a result, the implications are vast
  - Those affected by covid-19
    - During the pandemic, there was a lot of A.I generated content that led to distrust in public healthcare
  - Venezuelan people
    - The Venezuelan state media used A.I generated videos to spread pro government propaganda
  - American voters
    - I personally have seen numerous A.I generated videos of Trump and Biden.
    - Usually, I'm able to tell whether the video is fake based on the voice sounding A.I like or the content being comedic, but a lot of people in the comments believe these videos are real

What needs to be done by who

- Social media companies need to devote more resources into detection of fake news.
  - Different platforms could work together to develop an efficient and accurate method to counter LLM Based attacks.
- Certain measures such as community notes are already in place but these require human interaction and these can be faked.
  - Many people on social media do not look further to fact check.
- Companies need to devote more time for spam detection as bot accounts can run rampant.
- Users on social need to spread awareness about fact checking.
- Self education on fact checking important information.

Code Demo

https://colab.research.google.com/drive/1ESd0VO6Gi4FTkOGgMqQo51nqDkFev0gV?usp=sharing

Sources

- https://arxiv.org/pdf/2310.10830.pdf
- https://www.aoml.noaa.gov/general/lib/lib1/nhclib/Fake-News-WorksheetProQuest.pdf
- https://usa.kaspersky.com/resource-center/preemptive-safety/how-to-identify-fake-news
- https://www.brookings.edu/articles/how-to-combat-fake-news-and-disinformation/