The good, the bad and the ugly of using AI tools

By Paul Morgan
15 Mar 2023

I'm blown away by the sheer human-like quality of the language model that has been produced by OpenAI like most people are, it's an amazing product. We at Altron Karabina are experimenting with using the tool to verify and optimise code, such as SQL-stored procedures and Python scripts. Salespeople are also using the tool to produce succinct summaries of complex products to help clients understand them better.

Paul Morgan, head: data analytics and AI at Altron Karabina | image supplied

The good

The promise of AI tools is that they can supplement human endeavour - read *Human + AI* by Paul Daugherty and James Wilson for more on the subject. A lawyer, financial adviser, marketing manager or programmer who can use ChatGPT, LaMDA or other AI tools effectively is going to be more productive, and possibly more inventive than one that can't. AI technology has the potential to supercharge human work, and there is no doubt that the potential is there to create more jobs and more prosperity, not less.

The bad

However, technology such as ChatGPT, like all human inventions, is a double-edged sword (or ploughshare) and can be used for good or ill. ChatGPT declares these concerns itself in almost every chat we've had.

For those who remember, in 2019 OpenAI decided to withhold the release of GPT-2, a previous version of their language model, citing concerns about the tool, which could generate convincing news articles, being too easy to use for misinformation purposes. Given that the following years brought us the attack on Capitol Hill and Covid-19 itself, ripe topics for misinformation, you can't help but wonder if OpenAI was using its own prescient tech to predict the future.

There are a number of issues – firstly, the tech itself can produce misleading, wrong or damaging information. In one reported incident that I've been unable to verify, a GPT-3 application (not ChatGPT) allegedly produced the following responses to a mock patient with suicidal tendencies.

Whether this exchange actually happened or not, there are plenty of other verified exchanges that show that AI models can be impacted negatively by the content that they learn from - see Microsoft's twitter-fed Tay chatbot launched in 2016, and live for 16 hours before it was rugby-tackled off the field for generating offensive tweets.

OpenAI has put in a lot of effort to try and avoid the negative effects of training language models on a large corpus of documents, such as Twitter, Wikipedia, and various sections of the internet, by having humans tag text that is violent, racist, misogynistic or otherwise unacceptable, to ensure that it doesn't contaminate the user experience with GPT-3.

It's good to see that organisations like OpenAI can be self-regulating in this regard, but can we be sure that all AI businesses will do the same? The EU is preparing regulation that might help with this, the [AI Act](), expected to be made law in 2023, but the impact remains to be seen.

The ugly

The other concern is the damage that can be done in the process of creating the technology itself. As I write, Microsoft, OpenAI and GitHub are defending a class action that alleges that the corpus of programming code that was used to train GPT-3 contains licensed code that should not be profited from. Matthew Butterick, who filed the suit, states that the creation of CoPilot, GitHub's GPT-3 based coding assistant, is "software piracy on an unprecedented scale".

I've started wondering if we are not breaching any copyright laws by using ChatGPT to check my team's code against someone else's code, albeit unknowingly. As a throwback to the previous concern, the coding Q&A site, Stack Overflow, has banned AI-generated answers to programming questions, saying "these have a high rate of being incorrect". This concern must also apply to the millions of non-coding articles and books that GPT-3 has been trained on – who really owns the poems, scripts and movie scripts generated by ChatGPT?"

Generative AI like ChatGPT reveal systemic issues beyond the tech industry
Jordan Richard Schoenherr< 6 Mar 2023

In addition, Time reported in January 2023 that the very act of labelling some of the internet's most offensive content to reduce any potential toxic output from GPT-3, has also caused damage. This work is generally outsourced to countries where labour is cheap, and a Kenyan company that paid workers less than $2 an hour for the task of labelling troubling material, seems to have had issues with employees who allege emotional trauma dealing with the nature of the content they vetted.

OpenAI is by no means the only AI company that uses low-cost labour for this purpose – last year, Time published another story about the same Kenyan company performing labelling work for Meta in the article Inside Facebook's African Sweatshop.

The verdict

The impact of AI is far-reaching – like many human inventions since fire itself – and we will have to guard it carefully to ensure that it keeps us warm, rather than burning out of control.

The legislation will slowly appear that will help to curb some of these issues with AI tools, but in the meantime, if you're deploying AI solutions in your business, it may be worth adding an additional ethics gate in your development process to debate these risks before release.

ABOUT THE AUTHOR

Paul Morgan, head: data analytics and AI at Altron Karabina.

For more, visit: https://www.bizcommunity.com