ChatGPT の性能確認

中居駿太　金子藍璃　青木実乃利　門谷志音　白石将大

発表要旨

AI 技術の問題の一つにハルシネーション（事実にない情報をあたかも本当かのようにでっちあげること）があります。これは将来、名誉毀損や情報漏洩を引き起こす可能性のある非常に大きな問題です。今回私たちは ChatGPT と Python を用いて、ハルシネーションの進行具合や特徴を研究しました。その結果、ハルシネーションが実際に進行しているデータや、特定の分野において、質問に対する AI の正答率の低下を示す興味深いデータを得ることができました。

１．背景と目的

私たちは、急速に発達する AI 技術に問題はないのかという疑問を持ちました。そんな中、「AI は事実に無い情報をあたかも本当のことかのようにでっちあげることがある」というハルシネーションの問題を取り上げた記事を見つけ、その情報の真偽や実態を調査したいと思い研究に踏み切りました。

２．方法

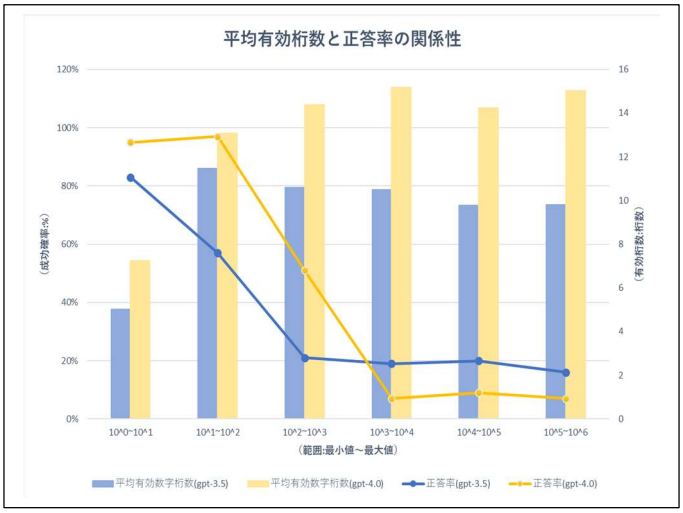
ChatGPT の GPT-3.5 と GPT-4.0 を使い、それぞれに Python を用いて四則演算をするように指示をします。そして回答結果の正確性や、ミスの仕方などのデータを集め、GPT-3.5 と GPT-4.0 とでデータを比較して、ハルシネーションの進行や特徴を調査しました。

３．結果

（１）GPT-4.0 は GPT-3.5 に比べて、四則演算を正確性 80%以上で回答できる桁数が二倍以上になっている。

（２）GPT-3.5 では、わからない問題に対して 4 回に 1 回程度は「わからない」と回答をしたのに対し、GPT-4.0 ではどんなに問題が難しくなろうとも絶対に「わからない」とは言わず、間違っても回答をする。

（３）割り算の回答における有効桁数と正確性を調べたところ、GPT-4.0 は GPT-3.5 に比べ有効桁数は多いものの、正答率は下がっていた。（右図）



４．結論

（１）GPT-4.0 は GPT-3.5 に比べて、正確性の面で性能が大きく向上している。

（２）ハルシネーション（でっちあげ）が進行している。

（３）GPT-4.0 は能力以上のことを妥協できずに行うため、失敗が増えている。

５．今後の課題

1 つ目は、ChatGPT の回答がどのような情報をもとに作られているのかを調査し、間違った回答をする理由を突き止めることです。2 つ目は、ChatGPT に計算に強いモジュールを入れたときの正確性や傾向を調べることです。

６．参考文献

・ChatGPT の可能性と危険性ハルシネーション問題/著:児玉龍彦

・https://atmarkit.itmedia.co.jp/ait/articles/2303/30/news027.html

７．IT・データサイエンスの活用

今回の研究では、質問の機構、質問をわかりやすい形に処理する機構、傾向を調べる機構を自動化した「Python のプログラム」で作成しました。Python を活用することで、数万にも渡るデータ収集が可能になり、より誤差の少ない分析を手早く行うことができました。