



生成 AI 時代をどう生きるか？

神戸大学 経済経営研究所
講師 早木 祥夏

今年の4月からクロスアポイントメントで研究所に所属しています。普段は高松で研究しておりますので、どうせなら気の利いたうどんの記事でも、と言いたいところですが、今回は最近の流行りでもある生成 AI についてコラムを書かせて頂ければと思います。私自身は特に生成 AI の仕組みに詳しいというわけではないのですが、先日、生成 AI と研究倫理について社会科学系研究者の立場から考える機会がありましたので、そこでの知見を簡単に共有いたします。

生成 AI は ChatGPT に代表されるような対話型の自然言語処理システムであり、すでに使ってみたことがあるという方も多いと思います。折角なので ChatGPT 3.5 に「生成 AI とうどんの共通点は何かありますか？」と尋ねてみました。出てきた答えの中で興味深いものがありましたので、以下に抜粋します¹。

生成 AI の動作はモデルとアルゴリズムに基づいており、トレーニングデータに影響を受けます。一方、うどんの味やスタイルは地域や伝統に基づいており、地域の文化や歴史に影響を受けます。両者とも創造の背後には特定の要素や文脈が存在します。

ややこじつけにも思えますが、生成 AI におけるトレーニングデータの重要性がよくわかる内容になっているかと思えます。香川県のうどんが比較的スタンダードなものである一方、群馬県の「ひもかわうどん」なんかは非常に特徴的です。文化的背景によって、うどんという出力が大きく変わりうるということになります。生成 AI の話に戻りますが、その問題として「間違った答えを返す」ということがよく取り上げられます。実際、トレーニングデータにバイアスがあると、生成 AI の出力にもバイアスが生じると言われています。

興味深いことに、人間の意思決定における様々なバイアスが生成 AI にも発生することがあると言われています (Rich/Gureckis [2019])。そのようなバイアスの一例としては、結論ありきでそれを追認するようなデータばかり集めてしまうという確認バイアスが挙げられます。差別や陰謀論の醸成につながりかねない危険なバイアスではありますが、人間だれし

¹ 正確には ChatGPT 3.5 に「生成 AI とうどんの共通点は何かありますか？」と打ち込んだ後、「もっと面白い切り口はありませんか？」と打ち込んだ結果の抜粋です (2023/11/08)。このように対話を繰り返しながら、答えを微修正したりアレンジしたりできるところが ChatGPT の重要な利点の一つでもあります。

も多少は持っているバイアスかと思います。生成 AI でも似たようなことが起こってしまうというのはなんとも人間臭く、逆に親しみが湧いてきませんか？もちろん、間違っていたりバイアスがあったりするような答えを返すというのは重大な問題であり、その点にどのように対処するか真剣に議論しなければなりません。

生成 AI と学術研究の付き合い方については、Van Dis et al. [2023] や Hosseini et al. [2023] で詳しく議論されています。詳細は省きますが、共通の結論として、生成 AI の出力の正確性を判断するには、その分野の専門家が注意深く確認する必要があると主張しています²。当たり前と言えば当たり前なのですが、だからこそ重要な論点かと思います。研究者は、特定の分野の専門家として、生成 AI の誤りやバイアスを指摘する最後の門番となりえます。そう考えると、生成 AI 時代において専門家としての研究者の役割はより重要になっていくと言えるでしょう。

以上から、私も研究者の末輩として、ひとまず生成 AI に完全に仕事を奪われることはなさそうで一安心したいところですが、この手の分野の成長は日進月歩、秒進分歩です。みなさんがこの原稿を読まれる頃ですら、どうなっているか予想もつきません。近い将来、論文のアイデアを練るところから分析や執筆に至るまで、論文作成のすべてのプロセスを生成 AI が代替可能になる可能性も指摘されています。生成 AI の能力がどのように向上していくのか、それに対してどのように向き合っていけば良いのか、月並みですが、常に考え続けることが重要かと思います。

参考文献

- Hosseini, M., Rasmussen, L. M., & Resnik, D. B. (2023). Using AI to write scholarly publications. *Accountability in Research*, 1-9.
- Rich, A. S., & Gureckis, T. M. (2019). Lessons for artificial intelligence from the study of natural stupidity. *Nature Machine Intelligence*, 1(4), 174-180.
- Van Dis, E. A., Bollen, J., Zuidema, W., van Rooij, R., & Bockting, C. L. (2023). ChatGPT: five priorities for research. *Nature*, 614(7947), 224-226.

² 特に、Hosseini et al. [2023] は、論文の執筆プロセスにおいて生成 AI を利用した場合は、使用したシステム、日時、プロンプトなどを論文中に明示すべきとしています。これらを明示することで、査読者に該当箇所をより注意深く確認するよう促すことができます。なお、この原稿でも脚注 1 でこれらを明示しています。