



ChatGPT と漱石の本籍地

神戸大学 経済経営研究所
准教授 藤村 聡

生成 AI の進歩は著しく、すっかり日常生活に溶け込んでいる。先日に行った大学の授業ではレポート作成に生成 AI の利用を推奨し、その際のアンケートによれば週 2〜3 回ほど生成 AI にアクセスする学生が 5 割、毎日アクセスするハードユーザーは 2 割、今まで使っていなかったが便利なので今後は活用したいという学生が 3 割であった。いずれも利用を始めたのはここ 1 年くらいであり、まさしく急速に普及したと言えるだろう。しかし利用が進むにつれ、その限界や稚拙さを指摘するコメントが増えてきた印象はぬぐえない。今日までに複数の対話型の生成 AI が公開されており、ここではその代表的存在というべきオープン AI 社の ChatGPT（無料版。以下 GPT と略記）の実情を見てみよう。

GPT が提供する情報は多彩である。たとえば画家ゴッホについて、美術界の評価を受けることなく他界したという世上の理解にはずっと疑念を抱いていた。死去前の画風はそれまでの単調な陰鬱さを脱却し、前人未踏の独創性あふれる境地を切り開いたようにしか見えないのである。その疑問に GPT は画家の友人や美術評論家から死の前年には評価されていたと回答し、根拠となる評論も明示する。このほかにお薦めの温泉やヘルスケアの助言、はては姓名字画の運勢占いなど GPT がカバーする領域はほとんど無限である。

しかし思わず首をひねりたくなるような誤った情報を平然と主張するケースがあることは否めない。最近に遭遇した事例を紹介しよう（2025 年 5 月時点）。文豪夏目漱石の本籍が北海道の何処であったのか、具体的な住所を調べる必要に迫られた。東京生まれのちゃきちゃきの江戸っ子で、松山や熊本あるいはロンドンに短期的に居住したものの、ずっと東京に住居があった漱石は、実は帝大生時代から死去の数年前まで本籍を東京ではなく、北海道に移していたことはあまり知られていない。その理由は漱石本人が沈黙しているので明確ではなく、徴兵逃れではないかと漠然と推測されるに留まる。

ともあれ漱石が本籍を北海道の岩内町に置いていたことは公的な本籍簿や各種の文献で疑う余地がなく、岩内町にはその記念碑も設置されているという。そこで、まずは「漱石の本籍地は北海道か？」と問うたところ、GPT は以下のように回答した。

「いいえ夏目漱石の本籍地は北海道ではありません。彼の本籍地は一般的に『東京都新宿区（旧・江戸牛込）』とされています。（中略）北海道が漱石の本籍地とされることは、史実的に確認されていませんし、そうした記録も残っていません。もしかすると、誰か別の人物と混同されているのかもしれませんが、確認してみましようか？」

説得力にあふれる語調であり、このまま信じてしまいそうである。気を取り直して「夏目漱石の本籍は北海道か？」と問い掛けると、

「夏目漱石（本名：夏目金之助）の本籍は北海道ではありません。彼の出生地は江戸（現在の東京都新宿区）であり、本籍も東京都内にあったとされています。北海道に本籍があるという情報は、誤解やデマである可能性が高いです。夏目漱石が北海道に特別な縁があったという記録も基本的には存在していません」

誤解やデマという厳しい表現が使われ、倦むことを知らない GPT もしつこい質問にはキレるのかもしれないという懸念を抱いた。

見極めるべくは、こうしたミスが発生する理由である。インターネットに散らばる情報を拾い集めて整理し、それを提供するのが GPT の基本的な機能と言えるが、その際には AI アライメントと呼ばれる人間のチェックが入っている。これは初期の対話型 AI が悪意のあるユーザーに影響され、人種差別的な発言を繰り返すようになった事故を回避するための措置と言われており、同時に情報の精度向上に寄与していると思われる。とはいえネットに流れる情報は莫大であり、GPT が拾いきれない（整理しきれない）不備が出るのはやむを得ない。GPT 自身も画面上で、重要な情報は自分で確認するように注意喚起する。

GPT の誤回答に系統的な規則性があるか否かという点にも興味をひかれる。もし GPT が苦手になっている領域が存在するならば、それは生成 AI の特性を考えるうえで興味深い現象である。すでにネコ画像の判別と作成をクリアした AI 研究のさらなる課題かもしれないが、残念ながらオープン AI 社は情報を公開していない。

現在のところ、インターネットにおける情報の信頼性はウィキペディア、グーグル検索、そして GPT という順序のように思われる。ウィキペディアは博識な利用者が記述を修正して他の読み手が再び検証することが可能であり、グーグル検索は自分で複数の異なる情報を判定できる。GPT の情報提供は一方的で不可逆である（SNS？ 論外である）。

しかし本稿で指摘した GPT のミスもやがて修正され、さらに精度を高めていくのは間違いない。その一方で、個人的には当初から没個性的な AI に独創性や新しい価値観の創出は期待しておらず、映画「ターミネーター」や「マトリックス」で描かれた人間の能力を遙かに凌駕し、全知全能的な AI が登場するのは彼方の未来か、または永久に実現は不可能ではないかという気もする。理由は簡単である。AI が依拠する知識の基盤は欠陥だらけの人間が作った情報であり、その不完全さは歴然としているからである。

とりあえず現在の我々が気に留めるべきは、次の警句ではないか（出典は不明。ロボットを題材に傑作 SF 小説を発表した作家のアイザック・アシモフ、あるいは様々な作品で未来世界を活写した手塚治虫や星新一かもしれない。GPT に訊いても明確な回答は得られなかった）。文中のロボットという言葉は、そのまま AI に置き換えてかまわない。

「私はロボットが人間に近づくのは怖くない。人間がロボットに近づくのが怖い」