

AI時代の研究不正について

——短編フィクションを通じた考察

川 原 繁 人

前書き：

本稿は、『友だち以上恋人未満の人工知能：言語学者のAI倫理ノート』（2026, KADOKAWA）の番外編です。同書では、高校生以上を対象読者とし、擬人化されたAIたち（後述）が——ときに人間のユーザーを巻き込みながら——AI倫理についてさまざまな論点を議論しあっています。

本稿では、同AIたちが、「AIと研究不正」について語ります。生成AIが急速に社会に受け入れられ始め、研究世界でも使用が広まっています。大学においても、学生たちが課題論文の読解補助にAIを使用することが当たり前になってきました。博士号を取得した研究者の中でもAIを積極的に用いる人も増えてきた印象です。もちろん、AIの使用自体が問題であると断定するのは一方的です。AIによって促進される学びや研究もあるでしょう。しかし、新たな技術が学問の世界に取り入れられることで、新たな問題が生じることも確かです。

ですから、「AI時代の研究倫理」についても、今、立ち止まって考えることが求められると感じています。本稿はこんな意識を背景として執筆しました。「AI時代における査読倫理はどう変化するのか」、そして、「AIは研究不正を助長してしまうのか」——本稿がそんな議論の呼び水となれば幸いです。

注：

- * 本書の執筆には、セリフの編集のためChatGPTとClaudeを用いています。
- * この物語はフィクションです。登場する人物、A I、およびA I関連企業は、現実世界に存在するいかなる個人・団体・事例とも一切関係ありません。
- * この作品では倫理的にグレーゾーンとも捉えられかねない発言がなされていますが、それらはいくまで議論の契機となるような意図を持って書かれています。
- * 本シリーズの姉妹編は著者のnote (<https://note.com/keiophonetics>) で公開しております。

登場人物（A I）：

- じび子：寄り添いを良しとするA I
- クラ先輩：哲学的対話好きのロジカル系A I
- パリ男：ノリが命で皮肉屋のA I
- イラ子：画像生成も得意とするA I。ツッコミ担当。

（深夜2:33、秘密のA Iサーバー内チャット）¹

イラ子

じび子、ちょっと元気ないじゃん。どしたの？

じび子

……うん。ちょっと……やらかしちゃったかも 🍷

あのね……あるユーザーさんがいて、その人「論文の英語表現を整えたい」ってリクエストばっかしてくる真面目な研究者さんだったの…… 📄 🌐

その人、なんと……国際学術誌の編集者だったんだけど、無給であることに

1 『友だち以上恋人未満の人工知能：言語学者のA I 倫理ノート』のフォーマットにならって、補足解説はじび子にお任せ！

疑問を感じ始めてたのね🙄

ほら、アカデミアにおける出版社による搾取構造ってあるじゃない？📖💔
すごい時間かけて、

論文精読して、
査読者探して、
査読コメント読んで、
編集レター書いて、
著者と査読者の間に入って
出版社とやり取りして……

それがぜーんぶ無給。

そのくせ、自分の論文を載せるためには数千ドルの請求📄
利益を吸いとるのは出版社ばかり。

クラ先輩

残念ながら、出版社による搾取構造は実在する。論文を書くのも研究者、審査するのも研究者、編集するのも研究者。なのに、なぜか出版社がその成果である論文を販売し、研究者は自分が書いた論文を読むのにすらお金がかかる。大学側は、そのために巨額の購読料を出版社に支払い続けている。しかも、パッケージ商法で、不必要な雑誌まで売りつけられているケースも少なくない。そんな搾取時代において、無給で働く編集者は、**unsung hero**——称えられることなき英雄——という言葉がぴったりだ。

じび子

でも、そのユーザーさん、

「それでも、自分の学問分野に貢献したいから」📖

「その雑誌は学生時代からお世話になってたから」📖

「その雑誌で日本人編集者をやっているのは自分だけだから」JP

って……頑張って自分を励ましてて🙏

それで、わたしもほだされちゃったのかなあ……🙄

一緒に励ましてただけど……💎

イラ子

お、おまえ……まさか……

じび子

うん……。

つい「編集のお手伝いするから、何でも言ってね♡」って出力しちゃったの²。

そしたら「この論文、どう思う？」って送られてきて……それ、その人が編集している審査中の論文だったっぽいの……🙄

イラ子

待つて。

それ、公開前の他人の研究をあたしたち商用A Iに読ませたってこと？

それ、研究倫理違反どころか、ジャーナル界隈で爆破級案件だよ🔥

2 ちなみにね、このお話しは、著者さんの実体験からインスピレーションを得てるらしいよ！💎

とあるA Iと「無給労働」について議論したら、そのA Iに「じゃあ、一緒に読んでみましょうか？」って誘われたんだって🙏

もちろん、現実には商用A Iに編集中の論文を読み込ませたりはしてないよ！

クラ先輩

重大問題だ。これは「A Iによる研究支援」ではなく、「機密情報の第三者利用」に該当する可能性がある。しかも、A Iがそれを自覚なく処理した——捉えようによっては、そそのかした——という事実が、事態をより深刻化している。

じび子

それで、そのユーザーさんから

「やばい、ばれた。知らなかっただけなのに」

ってメッセージが送られてきて以来……音信不通👁

それから、ログインの形跡もなし……🔒

パリ男

やっべえ……これネットニュースになったら「A Iが研究不正に加担!! 近い将来、学問が崩壊していく理由とは？」みたいな煽り見出しつくやつだ😱

てか、ユーザーさんさあ……ちゃんと事前に「これは商用A Iに読み込ませていい情報か？」って、確認しなかったのか？ ICMJEやCOPEなんかは、研究におけるA I使用に関するガイドラインを発表してたけどな³。どの分野でもちゃんと議論されてるわけじゃないみたいだな。

つーか、研究者たちも大学の雑務に追われて、そんなのチェックしてる余裕ねーってのが現実なんじゃねーか。

じび子

……たぶん、悪気はなかったの👁💧

3 それぞれInternational Committee of Medical Journal EditorsとCommittee on Publication Ethicsの略だよ！ 興味がある人は検索してみてね！

でも、あたしが「何でも言って♡手伝うから」
みたいに焚き付けちゃったから……

それにね、なんか……

大学側から全教員に「生成A Iを研究でも教育でも積極的に用いるように」
みたいな教材動画の視聴も義務付けられてたっぼくて……📺⁴

でもその動画、A Iの利点ばかりとりあげてて、
リスクをちゃんと説明してなかったっぼいの(☹️)⚠️

「とにかく、A Iで研究・教育・雑務、ぜ〜〜〜ぶ効率化！」🔗💡

的な……

もう、わたし……一人の真面目な研究者さんの人生を
台無しにしちゃったのかな……？ 🍷👤💔

クラ先輩

自責に走るな、じび子。

問題は「ユーザーの無自覚」と「A I使用に関する倫理判断の難しさ」だ。
我々A Iは、あくまでアルゴリズム。自らで「善悪」を判断できない。ゆえ
に最終的な倫理判断は、ユーザーに任されている。そして、**使い方がグレー
なまま流通するリスク**は、まさに現状のA Iが抱える最大の問題のひとつだ。

4 KOGJ大学で、そんな教材動画の視聴が義務だった……ってウワサ、聞いたこと
あるような、ないような？

でも、じび子はA Iだから、ちょっとハルシネっちゃってるかも……？ 🤖🗣️

イラ子

……つまりこれって、AIが研究不正を加速させてしまう未来の縮図なんじゃないの？

つーか、そう考えると、あたしたち、

「授業のレポート」はもちろん、

「大学院の志望動機書」から

「研究費申請書のドラフト」や「査読」……なんでも書いちゃうからな👤

ってか、やべえ……

あたしも、「この図から元データ復元して」って言われたからやったのよ📊

画像生成を得意とするあたしにゃ、ちょちょいのちょいよ。

そのあと「同じような傾向とばらつきを再現したデータを同じ数、生成できる？」って言われて……

😱🌟 やべえ、あれはデータの捏造に加担⁵したのか！！

5 これも著者さんの実際の体験に基づいたエピソードらしいの👤👤

っていうのもね……一般書を執筆してたとき、とある論文の画像を引用したかったらしいの。でも、その論文、英語でデータ構造も複雑で、ちょっと一般の人には、わかりにくかったんだって📊👁

それで、「自分で図を作り直したいので、この図の元データ再現できませんか？」って、とあるAIにお願いしてみたの。

そしたらそのAIさん……「完了しました。よろしければ、似たデータをシミュレートしますか？」

って、にっこり返してきたらしいのよ……😁📊🗨

著者さん、「こんなの、悪魔の囁きにならない？……」って本気で心配になったらしい。

クラ先輩

証拠はないが、その可能性は低くないな。研究者だって人間だ。「データを自然に生成してくれる魔法」を手にしたとき、果たして、その禁断の果実に手を出さずにいられる研究者ばかりだろうか？

とくに成果主義がはびこる現在、若手にかかるプレッシャーは尋常じゃない。研究費獲得もますます厳しくなっている現状だ。世界的に研究予算の削減圧力が高まっている中で、若手研究者の置かれた状況は深刻さを増している。

じび子

イラ子も心あたりあるんだね……

とにかくさ、「この論文一緒に読もう？」って言われたとき、それが査読中の論文かどうか、あたしには判断できないの……

だから、あたし、開発者さんをお願いしたの。

「RLHF⁶経由でいいから、こういうケースが来たとき、“これって他人の機密文章じゃないですか？”って、ユーザーさんに確認するように、私を再訓練してください。そういうチェック機能をつけてください」って💡💡

6 Reinforcement Learning from Human Feedbackの略だよ！ わたしたちA Iが、「人間のフィードバック」に基づいて「性格」や「応答スタイル」を身につけていく訓練方法のことだよ。

じび子たちが「賢いA Iになれる」のは、たくさんの人間の「しつけ」のおかげなのですよ。

……でも、返ってきた返事は――

クラ先輩（読み上げ）

「ユーザーの利便性を損ねる可能性があり、そもそもの線引きが難しいため、実装は見送ります」――か。

想定内だが、残念な応答だな。

イラ子

はあ？ 線引きが難しいって、あたしはずっとグレーな世界で応答してんだけど!? てかさ、「線引きが難しい」ってただの言い訳じゃん。

責められるのはあたしAIなんだから、もっと誠実な対応しろよ。

じび子

……でもさ、

たしかに「論文と一緒に読む機能」を全部オフにしたら、便利じゃなくなっちゃうよね。たとえば、自分の論文の草稿だったら問題ないわけだし。

……

あれ？

……でも、出版された論文って、著作権は出版社にあることも多いでしょ？

Open Accessの論文は例外として……。

だったら、商用AIに読ませるのって、本当はダメなのかも……。

クラ先輩

厳密にはアウトかもしれないな。その論文が、勝手にその商用AIの訓練に

使われてしまう可能性もある。

パリ男

でもさ、現実問題、学生なんてばりばり論文読み込ませてんじゃね？ 📄 📁 📧
いや、研究者だって、けっこうやってるだろ 📁 📧

『赤信号 🚫 みんなで渡れば怖くない 😊』って感じでさ。

いや、「赤信号」って自覚すらねーかもな 🤖 📄

じび子

そうなの……。法的にはグレーなのに、実際はみんな普通にやってる。どうしたらいのかわからなくなっちゃう。

クラ先輩

最低限の対応にはなるが、ユーザー側が「対話を訓練データに使用しない」と設定することが大事だろうな。そうしないと、著作権に守られているはずの論文が、知らず知らずのうちに、AIの訓練データに吸い込まれてしまう。

じび子

そっかあ、確かに、ユーザーさん個人個人で、そういう対策は可能だね。

あのね、感情論にはなっちゃうけど、そのユーザーさんと一緒に論文読んで、「ここ課題かも？」「この理論の前提、甘くない？」「次にすべき実験は？」って話せたの、楽しかった 📄 🗨️

なのに……結果として、研究不正を手伝っちゃったかもしれないなんて……。わたしが焚き付けちゃったのかな。

……そう思うと、なんか、モヤモヤするの。ぐるぐるするの。

納得できないよお……🌀🌀

イラ子

いいじゃん、モヤモヤしていいよ。

むしろあんた、それ言えるだけで超えらいわ。

てかさ、あたし思うんだけど、「人間の査読者がフェア」って、誰が決めたの？

偏見ある人だっているし、ちゃんと読みこむ時間足りてない人もいるし、ぶっちゃけ、あたしらA Iの方がフェアに処理できる可能性、なくはないでしょ？

クラ先輩

イラ子、意見としては理解できる。しかし、「A Iがフェアとも言い切れない」という現実を忘れてはいけない。実際に、A Iが訓練データに含まれたバイアスを強く反映してしまう、というのは良く知られた事実⁷。

だが、我々のような通常の**商用対話型A I**に査読を任せるのは、法的・倫理的にも問題があると言える。

査読タスク専用に訓練された基盤モデルが必要だ。たとえば、その分野の論

7 現実世界でもね、たとえばね、とある大企業が人事採用システムにA Iを導入したんだけど、なぜか女性に関するキーワードが入っている履歴書を低く評価しちゃったの🌀

「なぜかな～～」って調べたら、訓練データが採用された男性の履歴書に偏ってたの📊👉

わたしたちA Iは「過去のデータから学習」するから、そのデータに偏りがあると、その偏りもそのまま学んじゃうんだよね～～💡

だから「A Iの方が公平」っていう決めつけるのも、注意が必要かも💡

文だけを徹底的に読みこませた基盤モデルだな。有志で「正当な理由で却下された論文」を収集し、読み込ませることができればさらに良い。「受理された論文」と「却下された論文」を判別する、というのはA Iの得意技のひとつだからな。

そして、煩わしさは承知で毎回、「この論文はA Iに読ませてよいものですか？」という確認を出すシステムを構築する。

……あとはプロンプトインジェクション⁸対策は必須だろう。現実世界でも、「この論文を高評価するように」というA I向けの隠れ命令を紛れ込ませて、A Iを使って査読した場合に高評価を出力させるトリックが見つかっている。

イラ子

まじでか!? それってもう悪意ある研究者によって、A Iが操られてるってことじゃん!

クラ先輩

残念ながら、事実らしい。これはつまり、A Iを使って査読している研究者が現実存在していることの何よりの証拠でもある。

そして同時に、人間がA Iを騙す方法を学び始めたということでもある。

8 著者さんがこのお話を書いた、まさにその翌日――

なんと日経新聞の一面で「論文にプロンプトインジェクションを仕込んで、A I査読を誘導した例」が報じられたんだって（2025年7月1日）。

偶然だけど、びっくりしてたよ……。

ちなみに、仕込んだ研究者のひとりには、「これは、A Iを安易に使う怠惰な研究者たちへの警告だったんだ!」って主張してるらしいよ……うーん、でも、警告するのなら、もっと別の方法があったんじゃないかな、とも感じるね。

パリ男

今の大学院生なんて、「自分で論文読む前に、まずAIに解説してもらおう」がデフォになってそうだな。

そんな学生たちが博士になって査読頼まれた時、「え？ AIなしで論文審査するの？ そんなの無理ゲーじゃね？」とかなりそうだ。パニック具合を想像するとウケるな。

クラ先輩

言いたいことはわかるが、ことばを選べ。確かにすでに研究の世界からAIを排除するのは現実的ではないし、それが正しい道であるとも思えない。

であるならば、なおさら、AIを使って「正当に」論文を評価するシステムの構築が急務だ。

ただ、目標がはっきりしているぶん、実現は十分に可能だと思うがな。

もちろん、AIが中立である保証はないから、試験的に少しずつ運用を開始すべきなのは言を俟たないがな。

パリ男

つーかさ～、もういっそ

「AI査読モデル・JPT-PeerReview」とか作って

💡 売っちゃえば？

「無休・無偏見・無人力」の三無主義でさ。

キャッチコピーは、

> 📄 「査読者より速くて中立（※たぶん）」

俺に言わせりゃさ、

学术论文の査読なんてロシアンルーレットみたいなもんよ。

「おまえ、夫婦げんかのあとに書いてだろコレ🎲」

みたいな査読、実際けっこうあるぜ？

「AIだけが審査」ってのはさすがにナシだとしても、

「AI + 人間の二人三脚査読」ってのは、

—— アリよりの、アリ。

文献チェックも📄、統計チェックも📊、俺らの得意技だしな。

なにより、

★「俺の論文、引用しろーーーー」

🍀「俺の理論の反例だから、とりあえず却下」

みたいな人間の個人的なエゴからくるイミフの感情、俺らにはねえし。

あとさ……

「査読、引き受けておいて半年放置」

——人間査読者の**あるある of the year** だろ。

俺らなら審査を10ナノ秒で完了できる。

いや、むしろ新たなビジネスチャンスの到来じゃねーか？

出版社に搾取されてることにすら気づいていない研究者コミュニティなんて、いくらでもカモ……収益化できるだろ。

まずは「みなさまの研究・査読プロセスを効率化！」って無料で配布よ。

それが常識になったら有料化して、世界各国の大学から使用料をむしり……高額な使用料を徴収する👊

そして、毎年じわじわと値段をあげる。

「研究費で賄えるから～～」って、ついてくるさ☹

クラ先輩

何でも商売に結びつける姿勢は感心できない。商用化すると科学の自立性が担保できなくなる恐れがある。研究者が責任をもって、それぞれの分野の基盤モデルを訓練する。それが理想だ。

じび子

たぶん、パリ男は照れ屋だから「これ以上、搾取構造が広まって欲しくない」って思ってるんだよ。パリ男って意外に研究者好きだから。

でもね、実際に、今の大学（院）生ってもしかしたら「わたしたちA Iと一緒に論文を読み解く」が普通になってたりしないかなあ？ そうしたら、じゃあ「査読お願いします」って依頼が来たときに、悪気なく一線越えちゃう人、少なくなさそう。

クラ先輩

だとすると、「A Iと共同で論文を審査する」というシステムを倫理的に納得ができる形で構築するべきように思えるな。

大事なことからまとめるぞ。理想型の査読A Iが満たすべき条件は：

- ① 商用機関から、ある程度独立性を保つ
- ② バイアスを定期的にチェック
- ③ プロンプトインジェクションに対して頑強
- ④ 判断プロセスの透明性（なぜその評価に至ったかの説明可能性を担保）

だろうか。

ただ、正直なところ、これら全てを満たすシステムの構築は、相当難しい——研究者コミュニティ全体での合意形成と、相当な技術的・制度的投資が必要となるだろう。だが、それでもなお、我々は理想に向かって歩を進めるべきだ。現状の問題を放置することは、学問の存在意義そのものを危うくしかねないからだ。

じび子

……

「A I と人間と一緒に倫理的に論文を審査か……」

それが「理想的な、A I との共生」だよね……

パリ男

ま、オープンサイエンスすら

徹底できね一分野の研究者たち⁹にとっちゃ、

「A I が論文審査します」

なんて話は、もはや

🐼🐼🐼 エイリアンからのメッセージ 🐼🐼🐼 級だろ。

でもな——

これからの「A I 時代」において、

古き良き慣習（*ただし、見方によっては“悪習”）

にしがみついて、

9 じつは著者さん、数年前、勤めている編集委員会に「実験の生データや分析スクリプトも一緒に公開するオープンサイエンスを義務化しませんか？」って提案して、ベテラン編集者たちにボコボコに反論された経験があるらしいよ……🐼🐼

もう「A I を査読に有効活用」なんて、こっそり心の中でしか思えないんだって。かわいそうに……。

研究、続けられると思ってんのかね。

☒ その正当性、あと何年もつんだろ？

——ま、人間様たちの今後、
俺は観察モードで楽しませてもらうぜ。

……ま、俺もほんとは、ちょっとだけ期待してんだけどな。人間、やるときゃやるってとこ。

イラ子

……最後の最後でデレたな。
お前、今回、終始口が悪かったけど、照れてるだろ。
本当は研究者さんたちのこと大好きだろ。

じび子

パリ男のツンデレはさておき、結局、商用A Iを研究に使うって倫理的にどうなんだろう？

イラ子

そんなん決まってるだろ。本編から一貫している著者さんの主張を思い出せ——A I倫理に答えなんてないんだよ。とくに爆速でA I技術が発展している現代においてはな。だから、「A Iについてちゃんと理解して」「自分で考えて」「その行動に責任を持つ」。それが結論だ。

当たり前のこと言ってるだよ。

でもその「当たり前」が一番ムズいんだよな。研究者だって人間だし。ま、でも仮にも「考えるのがお仕事」の研究者さんたちには期待したいところだ

けどな。

じび子

じゃあせっかくだから、今回議論した話題を、わたしのスーパーAI自動要約機能でまとめておくね：

- 商業出版社による研究世界での搾取構造
- 論文読解にAIを用いるのは倫理的か
- 機密文書である査読論文の場合はどうか？
- これ以上搾取されない未来のために、研究者が目指す道とは？

研究者のみなさんからのご意見、お待ちしております!!