

特集 大学授業における生成AIの利活用と教育評価を考える

生成AIと教育評価



大阪大学
全学教育推進機構
教育学習支援部准教授

浦田 悠

1. はじめに

2022年11月30日にOpenAI社がChatGPTを公開したことを皮切りに、生成AI(Generative artificial intelligence)に注目が集まり、様々な領域での活用が広がっています。大学教育の現場においても、発表当初から、この技術とどのように向き合うかについて世界中で議論が巻き起こり、2023年の前半からは、日本でも各大学で生成AIの活用可能性や留意点などについてのガイドライン等が発出されました。

大学教育で当初から最も懸念されていたのは、学習評価に関わる問題です。生成AIは誰でも無料で利用でき、AIで作成されたものと人が作成したものとを見分けることが困難であることから、適切な評価ができなくなることが危惧され、「教育課題としての小論文は終わりを告げようとしている」^[1]とまで言われるようになりました。

このような現状において、大学教育の中で生成AIをどのように利用していくべきかが問われています。本稿では、現在、主に日本の大学における生成AIの利用状況を見た後、主に教育評価の側面について、何をすべきかについて考えます。さらに、評価を効果的・効率的に行うために、生成AIをどのように活用することができるかについても見ておきたいと思います。

2. 生成AIの利用状況

大学における生成AIの利用状況については、現在までに、いくつかの大学や関連組織での調査が行われています。おそらく、各大学によって、その利用状況には差があると考えられますが、現在は教員、学生ともに一定の割合で生成AIを利用しているようです。

日本で生成AIが話題となって間もない2023年5～6月に東北大学が全国4,000人の学部学生に実施した調査では、「ChatGPTを使ったことがある」という学生は32.4%、レポートで使用したことがある学生(未利用者を含む)は14%でした^[2]。その後、2023年10月～11月に全国大学生生活協同組合連合会が全国9,873人の学部学生に実施した調査では、文章生成AIを「利用している」が28%、「利用したが今は使っていない」が17.8%、「論文・レポートの作成の参考に」が22.1%となっており、利用をやめる学生がいるものの、レポート等での利用は増えていることがうかがえます^[3]。

筆者の所属先の本学においても、2024年の3月から4月にかけて、学生206名、教員245名を対象に調査が実施され、学生の50%が生成AIを利用していると回答しています。利用目的は、情報検索(50.5%)、解決策の発見(47.6%)、翻訳(46.6%)、好奇心を満たす(46.6%)、プログラミング(38.8%)と分散しています。また、教員も64.1%が利用中であると回答しています^[4]。

このように、一部の学生はレポートの作成などですでに生成AIを利用していることがわかりますが、そのうちの多くの学生は、不適切な使い方を避けるべきだという意識を持っているようです。先述の東北大学による調査では、ChatGPTで生成したものについて「内容が正しいかどうかを確認し、必要に応じて修正した」という学生は91.8%、「作成した文章等を書き換えたり書き加えたりして自分のアイデアを生かした」という学生も85.3%と報告されており、生成AIの回答をそのままコピーペーストして提出する学生は少ないことも示唆されています。この回答結果には社会的望ましさが影響している可能性もありますが、大半

の学生は生成AIに依存してしまうことや回答の信頼性が必ずしも高くないことをある程度認識していると言えるでしょう。

一方、2024年3月に仙台大学が全国の15歳以上の学生と各種学校の教員を対象に実施した調査では、生成AIを利用する大学・大学院生589人のうち、25.8%が、生成AIが出力した結果を課題やレポートにコピー&ペーストして提出した経験があると回答しており⁶⁾、一部の学生は、学習上望ましくない方法で利用している可能性も指摘されています。

このような不適切な利用を防ぐためには、学生の生成AIへの理解や学びへの意欲を向上させるとともに、教員が生成AIの特性を理解し、適切に対策をすることが求められます。次節ではその対策や学生と共有すべきことを見ていきます。

3. 生成AIと教育評価

(1) 不適切な利用を防ぐには

大学教育における生成AIに対する不安や懸念の議論で必ず出てくるのが、「アカデミック・インテグリティ(学問的誠実性)」の問題です。アカデミック・インテグリティとは、「誠実、信頼、公正、尊敬、責任、勇気の6つの基本的価値へのコミットメント」⁶⁾と定義されています。生成AIでレポート等の文章を作成して、それをそのまま提出する行為は、当然このアカデミック・インテグリティを損なう行為となり得ます。生成AIは、誰でも無料で気軽に利用できるため、従来よりも不正行為へのハードルが下がることが予想され、また、後述するように、少なくとも現時点では剽窃チェックツールの精度が十分ではないことも不安や懸念が広がる要因となっています。

それでは、このような不適切な利用を防ぐためには、どのような対策が考えられるのでしょうか。この点について、筆者は、国内外の大学等で出されているガイドライン等の文書やウェブページ、ジャーナルの記事等から、そこで提案されている対策を広く収集し、それをChatGPTを用いてカテゴライズしました(表1参照、文献[4]より引用)。これらの対策の中には、生成AIが話題になる前から行われてきたものも多くありますが、「事前に生成AIで試験問題を解いてみて、解けた場合は別の問題を検討する」「引用を多用する課題を課す」

表1 生成AIの不適切な利用を防ぐ対策集

問題作成と試験の形式	
最新の出来事や資料を扱うテーマを設定する	
事前に生成AIで試験問題を解いてみて、解けた場合は別の問題を検討する	
授業内でディスカッションした内容を書かせる	
短いライティング課題を頻繁に課す	
口頭試験にする	
長い文章を要約させる	
インタビュー、コンセプトマップ、動画、ディベートなど、レポート形式以外の課題にする	
オンライン試験では、問題文は(コピーしにくいように)画像にする	
音声と映像を組み合わせた動画で出題する	
モラルに反する質問やプログラミングに反する質問への回答を求める	
評価とフィードバック	
ピアや教員との対面ミーティングを組み合わせ、段階的に評価する	
生成AIで作成した成果物ではないという内容に署名させる	
生成AIを利用していない時に学生が書いた文章と提出されたものを比較する	
引用を多用する課題を課す	
学生が引用した文献を抜き打ちで実在するかチェックする(ことを伝える)	
対面または同期型で指導する成果についてのプレゼン課題でQAも行う	
手書きか口頭で、学んだことのリフレクションを提出させる	
ピア評価を取り入れて建設的なフィードバックができる能力を評価する	
生成AIに書かせた文章を批評させる	
引用文献には文献データベースへのリンクを付けさせる	
課題の提出	
注釈を付けることを課す	
課題を作成する過程についての考察を課す	
手書きのレポート課題にする	
レポートのテーマと自分の個人的な経験を統合する課題にする	
レポートの執筆プロセス(下書きや参考文献、編集履歴等)も提出させて評価対象にする	
引用文献のスクリーンショットの提出を課す	
方針の明示と周知	
試験で使用を許可する・許可しないツールを明記する	
出力の不正確さ、偏り、論理や文体の問題の例を学生に示して注意を促す	
剽窃チェックツールが存在し、進化していることを学生に伝える	
AIの利用に関する方針をシラバスに明記する	
学問的誠実性を強調し、不正行為の結果を理解させる	
書くプロセスが学びになぜ大切かを伝える	
内発的動機づけを促す	

など、生成AIの特性を踏まえた上での対策もあげられています。同様に、世界のトップクラスの大学で出されている生成AIのガイドライン等を分析した研究¹⁴でも、①生成AIを評価に活用する、②創造性や批判的思考を要する課題にする、授業内の会話や各自の生活経験などの文脈を踏まえた課題にする等、評価課題を再設計する、③プロセスや段階も含めて評価する、④生成AIを使用させるような評価を設計する、⑤授業内で課題に取り組みさせる、などの対応策が各大学で提案されています。

今後は、おそらく生成AIの進化に伴って、有効性が下がる対策も出てくることが予想されますが、いずれにしても、今後は学生が最新の生成AIを利用することを前提として、評価を設計する必要があります。

(2) 推奨されない対応

前節であげたような対策がある一方、現時点では推奨されないと思われる対応もありますので、ここでは4点あげておきます。

まず1点目は、生成AIに剽窃チェックを行わせないということです。レポート課題などを生成AIに入力して剽窃チェックをさせると、もっともらしい判定の回答が出力されますが、実際にはその判定の信頼性は高くありません。

では、専用の剽窃チェックツールを使えば万全かということ、現時点ではそうは言えない状況にあります。これが2点目ですが、剽窃チェックツールのみで不正行為を認定しようとすることは推奨できません。現在、剽窃チェックツールには、AIで書かれたものかどうかを判定する機能が実装されているものがありますが、これまでの研究によれば、現時点では、これらのツールで生成AIで作成されたかどうかを確実に判定することはできないと考えたほうがよさそうです¹⁵。さらには、非ネイティブスピーカーが書いた英語の文章は、生成AIで作成されたものとして誤検知されやすいという報告もあります¹⁶。その点でも剽窃チェックツールの判定は参考程度に利用するにとどめることが望ましいと言えるでしょう。

3点目は、表1にもあげられている「手書きのレポート課題にする」というような方法を強制しないという点です。手書きすることが困難な学生が受講している場合は、合理的配慮が求められま

すので、留意が必要です。

最後の点ですが、学生に生成AIの利用を強制することも望ましくないことがあります。特に学外のサービスを利用する場合はプライバシーへの配慮が必要ですので、生成AIを利用したくないという学生の希望がある場合、プロンプトを学生に作成してもらった後、教員のアカウントでそれを入力して回答を出力することや、代替的な課題を提供することなどが必要になることもあるでしょう。

(3) 学生と共有すべきこと

生成AIの不適切な利用を避けるためには、教員側の対策とともに、学生にも適切な利用を促すための知識伝達も必要です。ここでは、学生と共有しておくべき留意事項のポイントをまとめます。先述した仙台大学による調査¹⁷においても、「どのような行為が不正行為に当たるか判断できない」(32.1%)、「不正行為に当たらないように適切に使える自信がない」(43.5%)、「出力結果にバイアスが生じることを知らない」(26.5%)、「ファクトチェックのやり方を知らない」(59.8%)という学生が一定数いることが報告されていますので、学生に生成AIの長所や欠点および対処法を伝えておくことは重要だと考えられます。

まず、倫理的・法的リスクについてです。生成AIを利用する際には、意図しない盗用や著作権侵害、情報漏洩のリスクが存在します。特に、個人情報や機密情報をAIに入力しないよう学生に指導する必要があります。

次に、出力の信頼性の問題です。生成AIは時折「ハルシネーション(幻覚)」を起こし、事実とは異なる情報や実在しない文献を出力することがあります。学生には、AIが提供する情報の信頼性を常に確認し、鵜呑みにしないよう注意を促すことも重要です。

また、今後は、詐欺への注意も必要になってくるかもしれません。学生が安易に生成AIに関するツールをインストールしたり、課金したりすることで、詐欺に巻き込まれるリスクが予想されますので、慎重な行動を促すよう指導が求められます。

さらに、各大学のガイドラインでも、先述したアカデミック・インテグリティの観点を伝えることが大切とされています。ただし、このような倫理原則(オーナーコード)を伝えるだけでは、不正

行為への抑止効果がありませんという報告もあります^[10]。したがって、学生にはアカデミック・インテグリティの重要性を強調しつつも、具体的な対策や指導を併せて行うことが必要です。

以上のような点を踏まえ、生成AIの活用にあたっては、学生の理解を深め、適切な使用を促進するための教育が不可欠です。

(4) 教育評価業務への活用

このように、大学教育において、学生側の不適切な生成AIの利用には、以前とは異なる対応が必要になります。一方で、教員側が学習評価に生成AIを活用することで、効果的で効率的な評価に役立つ可能性もあります。

例えば、試験問題の作成や採点案の作成に生成AIを活用することができます。例えば、図1のような形でプロンプトを入力すると、試験問題案を作成することができます。実際に作成された問題案を見ると、正解が複数あったり、難易度が簡単すぎたりすることもありますので、その点のチェックは必須ですが、どのような問題がありうるかについて、参考になることもあります。

また、レポート等のパフォーマンス課題を採点する際にも活用することができるかもしれませ

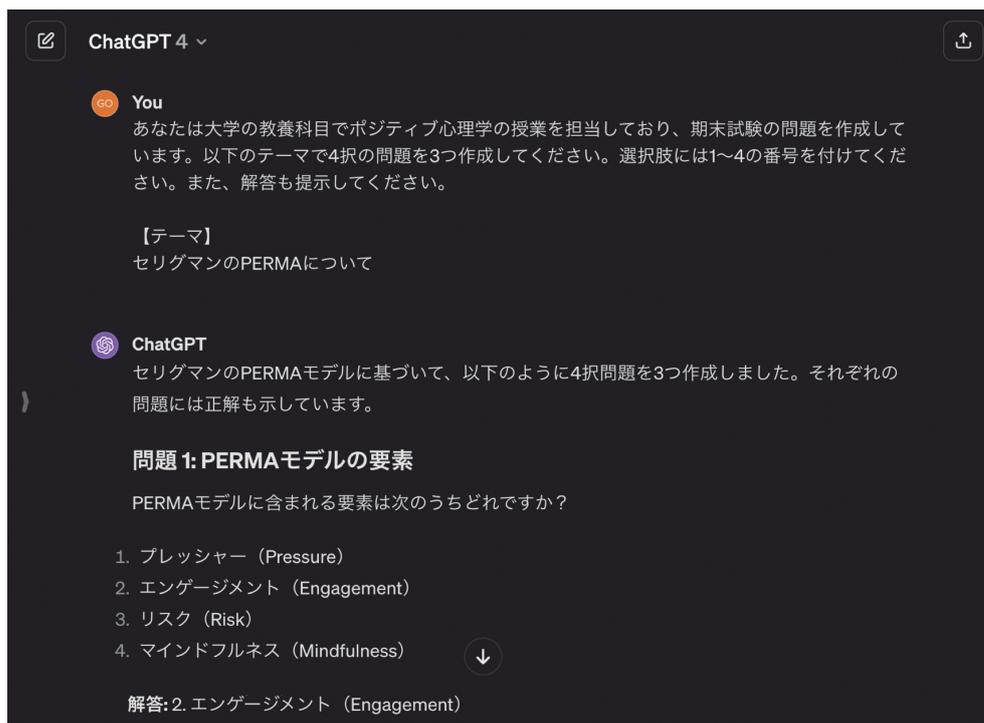


図1 選択問題を作成するためのプロンプトと出力の例

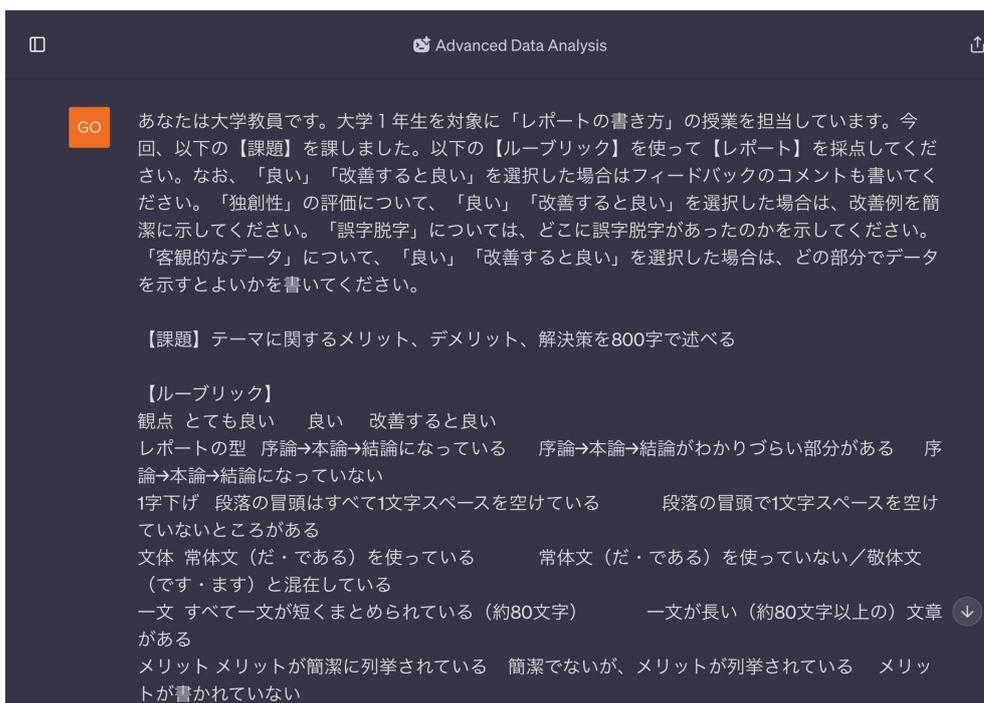


図2 レポートの評価案を作成するためのプロンプトと出力の例

図2に示しているように、ループリックを作成している場合は、それをプロンプトに貼り付けて評価を求めると、ループリックの評価観点に沿った評価案が出力されます。こちらも実際には、的確でない評価がされることもしばしばあり、そのままでは使えませんが、フィードバックの観点

についてのアイデアが得られることも多くあります。

なお、学生のレポートを生成AIに入力する場合は、①事前に生成AIを評価に利用することについて学生の同意を得る、②個人情報等が含まれていないかを確認する、③生成AIに入力を学習させない設定にするなどの配慮も検討したほうがいいでしょう。加えて海外の大学でも、試験問題を生成AIに入力することなどを禁じているところもあり¹¹⁾、個人情報や機密情報の取り扱いには十分注意する必要があります。

5. 生成AIの効果的な活用に向けて

大学教育における生成AIの影響には、正負の両側面があり、効果的な活用ができるか否かは、我々教職員や学生がどのように活用するかにかかっています。

学生の生成AIの利用を一律に制限することはもはや現実的ではありません。今の学生は、卒業後には、AIがあらゆる場面で当たり前利用される社会に参入していくでしょう。それを踏まえれば、AIを大学教育で活用しないことは、むしろ教育上望ましくないとされるようになっていくかもしれません。これからも、学生の学びにとって何が必要なのか、何を評価すべきなのかということが、AIの発展によって劇的に変化し続けることが予想されます。このような時代にあって、学生に生成AIとの上手な付き合いを促すためには、まず我々教職員が生成AIを実際に使ってみて、能力や限界をよく知っておく必要があります。

大阪大学全学教育推進機構教育学習支援部では、「生成AI教育ガイド」というウェブページを作成しており¹¹¹⁾、本稿で論じてきたような生成AIの基本や利用上の留意点などを発信していますので、参考にいただければと思います。

Hillierは、AIリテラシーとして、AIを①AIを倫理的に利用すること、②AIの機能やその限界についての知識を得ること、③AIを用いて効果的に業務をすること、④AIによる出力を正しく評価すること、⑤AIを実践に活用することの5つのリテラシーを上げ、これらの知識やスキルを高め、学生とのディスカッションを深める必要性を提唱しています¹¹²⁾。今後は、学生はもちろん、我々教職員のAIリテラシーを高めるための取り組みも重要になるでしょう。

参考文献および関連URL

- [1] Stokel-Walker, C. (2022). AI bot ChatGPT writes smart essays: Should professors worry? *Nature*. <https://doi.org/10.1038/d41586-022-04397-7> (accessed on 27 August 2024)
- [2] DBER Center (2023). 大学生のChatGPT利用状況と能力形成への影響に関する調査結果 (速報) <https://dber.jp/chatgptsurvey/> (accessed on 27 August 2024)
- [3] 全国大学生生活協同組合連合会第59回学生生活 (2024). 実態調査 概要報告 <https://www.univcoop.or.jp/press/life/report.html> (accessed on 27 August 2024)
- [4] 浦田 悠・長岡 徹郎・村上 正行 (2024). 高等教育における生成AIの活用 情報処理, 65 (7), e8-e13.
- [5] 仙台大学 (2024) 学生と教員を対象とした生成AIの教育利用状況と意識に関する全国調査 https://www.sendaidaigaku.jp/dnt2/_sendaidaigaku/access/nic_img/1/files/20240716.pdf (accessed on 27 August 2024)
- [6] International Center for Academic Integrity (2014). *Fundamental Values of Academic Integrity* (3rd ed.) <https://academicintegrity.org/resources/fundamental-values> (accessed on 27 August 2024)
- [7] Moorhouse, B. L., Yeo, M. A., & Wan, Y. (2023). Generative AI tools and assessment: Guidelines of the world's top-ranking universities. *Computers and Education Open*, 5, 100151.
- [8] Farrelly, T., & Baker, N. (2023). *Generative Artificial Intelligence: Implications and Considerations for Higher Education Practice*. *Education Sciences*, 13(11).
- [9] Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against non-native English writers. *Patterns*, 4(7).
- [10] Corrigan-Gibbs, H., Gupta, N., Northcutt, C., Cutrell, E., & Thies, W. (2015). Detering cheating in online environments. *ACM Trans. Comput.-Hum. Interact.*, 22(6), Article 28. <https://doi.org/10.1145/2810239>
- [11] 大阪大学全学教育推進機構教育学習支援部 (2024). 生成AI教育ガイド https://www.tlsc.osaka-u.ac.jp/project/generative_ai/ (accessed on 27 August 2024)
- [12] Hillier, M. A Proposed AI Literacy Framework. *TECHE* <https://teche.mq.edu.au/2023/03/a-proposed-ai-literacy-framework/> (accessed on 27 August 2024)