

生成AI時代の大学における成績評価

獨協大学
外国語学部ドイツ語学科教授 木村 佐千子



1. はじめに

生成AIをはじめとするAI技術の進歩は急速です。今年7月頃から、大規模言語モデルの開発にコストがかかりすぎて利潤が出ていないことが広く指摘され、IT大手企業の株価が下がったりしていますが、だからといって開発が止まるわけではなく、特に日本では少子化で人手不足が深刻化しているため、今後も企業はAIをとり入れていくと思われま

す。そのような社会に出ていくことになる大学生が、生成AI利用のメリットや留意点、限界についてきちんと知ることは有益でしょう。したがって、大学の授業や課題のすべてで生成AIの使用を禁止するのではなく、むしろハルシネーション等に気をつけながらツールとして補助的に活用する経験を積んでもらうほうがよいのではないのでしょうか。本稿では、どのようにすれば学生が生成AIを使って学習しても正当に成績評価ができるのかということを考えてきたいと思

います。様々な調査がありますが、普段から生成AIを使っている大学生は3~4割^[1]、用途としては論文の要約や、レポート・論文の作成が多いのが現状です。しかし、提出されたレポート・論文が生成AIによって出力されたのかを確実に判定できるツールはなく、AIの生成結果には「ゆらぎ」があるなど原理的な問題で、これからも当面の間、正確に判別できるツールは開発されないでしょう。2023年度にオンライン提出の授業レポートを読んで生成AIの使用を疑ったケースがありました。まとめ方などに生成AIのクセが感じられ、当該学生は留学生で学期はじめの提出物と明らかに文体も違ったため、授業後にその学生に生成AIからのコピーではないかきいたところ、認めましたので、以後、自力で書くようにと伝えました。このケースでは本人が認めましたが、本人が認めなければ断定はできず、学生との関係性によっては生成AI利用を疑ったこと自体が問題視されるかも知れません。そこで、生成AIを学生が利用したかどうかを個別にきかなくとも公平に成績評価できるような課題・試験等を考えていく必要があると考えています。

なお、あまり表立っては言いにくいことですが、

大学・学生のレベルによってとるべき対応は異なります。筆者の勤務先大学では、2023年7月に実施した筆記試験問題をChatGPT-4に解かせたところ、学生の最高点よりChatGPT-4の得点の方が高くなりました。しかし、学生の能力や習熟度によっては、AIより学生の方が高得点となることもあるでしょう。今年の春、東京大学の入試問題をChatGPT-4に解かせ、合格最低点に達しなかったというニュースもありました。一方で、ChatGPT-4は医師国家試験や弁護士試験に上位の成績で合格する水準でもあるということです。分野による違いとも言えます。本稿では、中堅私大の文系学部での講義・演習・外国語科目を主な対象として考えていきます。実習科目や理系科目では違うところがあると思いますがご了承ください。

生成AI時代の成績評価に関する筆者の提案は、昨年の口頭発表、あるいは2024年2月に発行された『情報学研究』第13号掲載の論文から基本的には変わっておりません。すでにご覧くださった方もいらっしゃると思いますが、要点を書かせていただきます。そのあとで、2024年度春学期に新たに開設した授業での成績評価について紹介したいと思います。

2. 生成AI時代の課題設定と成績評価の2つの方向

生成AI対応方針を学生に向けて公開している大学も少なくありませんが、「授業担当教員の指示に従う」となっている要素があることも多く、各教員が各授業のシラバスで生成AI利用の可否を明記することは不可欠だと思います。生成AIを使うべきではないと思って使用を控えた学生と、生成AIを使った学生がいた場合に、特に成績評価に影響が出るのが、レポートや小論文、外国語の作文、文献の要約課題、筆記試験などだと思われます。

(1) 成績に関係する試験や提出物で生成AIを使わせないようにする方向

上述のように、2023年の時点で、筆者の作成した試験問題については、履修学生よりChatGPT-4の方が得点が高いという事態が起きました。このこ

[1] <https://www.datascientist.or.jp/news/n-pressrelease/post-2828/>によれば29%。

とから、画面の向こうの学生が生成AIを使っているかどうか分からないオンライン筆記試験や、電子機器使用可の対面試筆記試験で、学生の実力をはかることはできないと考えます^[2]。生成AIに丸投げするほうが得点が高くなりかねません。生成AIの影響を排除して成績評価するには、筆記試験は対面で、電子機器の使用を禁止して実施するのが確実だと思われる^[3]。

文献の要約課題は、学生の力をつける上で重要ですが、自宅で書く課題とすると、生成AIを使って要約を作成できるのはもちろん、インターネット上の多数の文献要約サイトなども利用でき、昨年度採点したところかなり不正が増えている印象でした。自力で要約に取り組んでもらう場合には、授業時間内に、短めの文献を対面で要約させる方法があるでしょう。自分の意見などを書く外国語作文も、以前は宿題にすることが多かったと思いますが、生成AIの影響を排除して成績評価するなら、電子機器の使用を禁止して授業内で書いてもらうことを考えなければなりません^[4]。

学生たち、特に熱心な学生は、正当に評価されることを望んでいます。筆者自身が担当する授業以外についても、課題によっては生成AIを不正に使った人がいたのではないかというような意見(苦情)を学生から聞くことがあります。学生の信頼に応えるためにも、しっかりと対策を行うほうがよいのではないのでしょうか。

(2) 生成AIの使用を認め、これまでより質の高い成果を生み出させる方向

これからの社会でますますAIが使われていくことを前提に、AIを補助ツールとして活用し、課題でより質の高い成果を生み出す方向に導くことも考えられます。その際には、AIリテラシー教育を行い、学生本人が主体となって課題に取り組むこと、部分的にツールとして活用するにとどめ、決してコピー&ペーストで提出しないことなどをよく納得してもらう必要があります。

例えば、レポートでは、AIを使って自分では思いつかないようなアイデアを得たり、レポートの構成案を複数出させて検討したり、誤字がないか添削させたり、気の利いたタイトルを考えさせたりといった活用法が考えられます。とはいえ、生成AIからのコピーを検出するアプリ等がないなか、コピーでは提出できないレポート課題にする方法はあるでしょうか^[4]。

現段階で考えられるのは、信頼のおける書籍を指定するなどして使わせ、引用ページ数を脚注に明記する指示を出し、それを教員が確認すると伝えるこ

とです。生成AIはインターネットのURLを参照先にあげることがあり、書籍をPDFなどで読み込めばその内容をとり入れて書きますが、まだ自動で書籍のページ数を含む脚注をつけはしません。少なくとも書籍の引用部分は学生が自分で書くこととなります。また、生成AIはハルシネーションで誤った情報や実在しない人物を出すことがあるため、レポートの記述に情報の誤りや架空の文献等があれば大幅減点すると伝え、学生にファクトチェックを徹底させるのもよいでしょう。ただし、レポート執筆に特化したアプリ等も出ているので、注意が必要です^[5]。

昨年の文部科学省の文書^[6]では、レポートに口頭試験を併用することが提案されていました。しかし、生成されたレポートの内容だけを頭に入れてくれば、短時間の口頭試験は何とか乗り切れると考えられるうえ、これまでのように学期末にレポートを提出させると、長期休業期間に入ってから口頭試験をおこなうことになり、学生は出席しにくく、特に人数の多い科目では教員の負担も大きくなります。

そこで、提案したいのが少人数のアクティブラーニング型授業で、学習プロセスをもみる方法です。レポートで評価することが望ましい科目であれば、学期はじめにレポート課題を提示します。そして、授業内で文献検索、調べ学習やグループワークをした後、中間発表を実施します。調べ学習の段階では、文献はもちろん自分で読んでもらいたいのですが、例えば文献のスクリーニングにAIによる要約を使えば、効率よく多くの文献にあたれます。また、中間発表の準備段階で、発表の構成をAIに相談したり、文章の添削をさせたりすることもできます。文章添削を頼むと、改善すべき点とその理由を示すほか、AIがかなりほめてくれます。学生のやる気の維持には役立つように思います。

そして、中間発表での質疑応答や教員からのフィードバックをもとに、さらにブラッシュアップし、学期の終わり頃にプレゼン・質疑応答してもらいます。その後、プレゼンの内容をもとにレポートを書いてもらうというかたちです。このようにすれば、学習プロセスを観察しながら段階的な指導ができます。そして、少人数で学生との信頼関係が築きやすいため不正は起こりにくいと思います。AI時代において、教員と学生との間の信頼関係は、ますます重要になると思います。

3. 「学科横断演習」の例

次に、2024年度春学期に筆者自身が担当したアクティブラーニング型授業の例を簡単に紹介したいと思います。外国語学部全4学科の1～4年生を対象とし、定員20名で「学科横断演習 生成AIについ

[2] 電子機器活用の力をも含めて成績評価するのでしたら別です。ただし、その場合、生成AIは有料プランのほうが精度が高い傾向にあるため、試験の結果が学生の経済力に影響されることになり得ます。

[3] 科目特性によっては、持込不可の口頭試験で評価することも考えられます。

[4] ChatGPTで作成したレポートと自作のレポートを比較させたり、ChatGPTの出力を学生が添削したりする課題を出す教員もおられますが、現在では学生の使う生成AIも多岐にわたっており、一つのAIで出力したものを他のAIで添削させる等の不正も考えられます。

[5] <https://note.com/sharakusato/n/ne927ca24af33>, <https://reportai.jp/>など

[6] https://www.mext.go.jp/b_menu/houdou/2023/mext_01260.html

て知ろう」という授業を行いました。この授業では、口頭発表2回(5分と8分)に加え、授業最終回に小テストを実施しました。1学期14回の授業内容は表1の通りで、成績評価は中間発表25%、学期末プレゼンテーション25%、小テスト25%、平常点25%としました。2回の口頭発表については、学生相互にも採点してもらい、教員採点と半々の評価としました。この授業では、①生成AIについて知ることだけでなく、②学習プロセス、③信頼できる文献の活用、④プレゼンテーションの基本の習得、⑤学科・学年の垣根を超えての共同作業も重視しました。

表1 「学科横断演習」授業内容

授業回	内容	詳細
1	導入、対話型生成AIとは	使ったこと、名前をきいたことのあるAIサービスを挙げる 生成AIの定義・仕組み、使用上の注意事項について説明
2	生成AIを大学での学習に使用するには	4種の生成AIを実際に使ってみる(調べもの、アイデア出し、翻訳、英語の単語帳生成、文章添削、論文要約、音声会話、相談)
3	AI関連文献の検索と読解	グループワークのテーマを考えるための日本語・英語文献の検索 自分が発表したいテーマの絞り込み
4	中間発表テーマ決め グループ分け・分担決め	英語学習(4人)、英会話学習(3人)、英語教育(4人)、画像生成(3人)、音楽生成(5人)の5グループに分かれ、分担決め
5	文献収集、グループワーク	発表に使う文献をひとり3点以上見つけて、発表準備
6	レジュメ作成、グループワーク	発表レジュメの作成(レジュメのサンプル提示) 中間発表の予行練習
7	中間発表	3グループの中間発表(ひとり5分)、採点用紙記入(ひとりずつ)
8	中間発表	2グループの中間発表、採点用紙・グループ別コメント用紙記入と提出
9	中間発表フィードバック	各グループからのコメント用紙と教員が添削したレジュメの返却 中間発表1位グループの発表
	学期末プレゼンに向けて	学期末プレゼンに向けての改善点をグループで話し合う
10	スライド資料の作りかた	グループ内で相談しながら、プレゼン資料を作成
11	効果的なプレゼンテーションの方法	Zoom接続テスト(スライドを画面共有してみる) 予行練習
12	学期末プレゼンテーション	2グループの発表(ひとり8分)
13	学期末プレゼンテーション	3グループの発表
14	学期末プレゼンの投票 小テスト アンケート記入 1位グループの発表	ひとり2票でグループ別に投票 生成AIも利用可の小テスト 無記名手書きのアンケート実施 学期末プレゼンテーション資料を教員が添削したものを返却

学期はじめに生成AIに関する基礎知識を伝え、主に学習に役立つ使い方を実習してもらった後、文献検索をして適切な文献が3点以上見つかったテーマで3~5人のグループに分かれて口頭発表をしてもらいました。授業内でグループワークや予行練習の時間もとりました。1回目の発表はA4サイズで1~2枚のレジュメ、2回目の発表ではスライドを用意し、どちらにも3点以上の参考文献を明記してもらいました。資料作成や発表ではひとりずつの担当箇所を明確にしてもらいましたが、例えばレジュメはグループでとりまとめてPDF化してLMSにアップロードする、グループ全体でまとまった発表とするなど、グループワークを重視しました。小テストは、電子機器使用不可の対面試験を考えていましたが、学生たちに、生成AIを使う授業なので試験も生成AIを使うものにしてほしいと言われて成程と思い、電子機器も授業資料もすべて使用可能、ただし他人の力は借りないという条件で行いました。小テストの記載内容^[7]や最終授業で実施したアンケートから、

学生たちはある程度のAIリテラシーを身につけ、この授業で重視したかった上述の5点について目標は達成できたと考えています。特にグループ発表関連の項目は、評価が10点満点で平均9点以上と、とても高くなっていました。現在では、発表スライドもAIで簡単に生成できますが、この授業ではAIに丸投げしたプレゼン資料などはなかったと考えています。発表内容も、文献の情報と学生自身が生成AIを使った経験を組み合わせ、オリジナリティのある内容が多く、学生たちのフレッシュな感性が光っていたと思います。

この授業では、前学部長との相談で定員を20名に設定できたからこそ、2回の口頭発表の時間を確保することができました。履修人数の多い科目では、生成AIの活用を認めつつ、学生一人ひとりの毎週の取り組みを成績評価するような授業運営は難しいと思います。私立大学では、経営面を考えなければならないため、全ての授業を少人数にすることが難しいのは言うまでもありません。そのため、生成AI時代においては、人数の多い科目では電子機器使用禁止による対面試験で各分野の基礎を問う、少人数授業ではレポートやプレゼンテーションの力を伸ばすといったメリハリをつけたかたちで成績評価をしていくのがよいのではないのでしょうか。

4. まとめ

生成AIが社会で普及する中、学生には生成AIの活用法を身につけてもらいたいと思いますが、AIに課題を丸投げしては学生の能力が向上しません。一方、学生の提出物が生成AIによって作成されたものかどうかを確実に判定できるツールはありません。そのため、①成績に関係する試験や提出物で生成AIを使わせないようにする方向と、②生成AIの使用を認めこれまでより質の高い成果を生み出させる方向に分けて成績評価を行うことを提案いたしました。前者では電子機器の使用を禁止し、対面で実施する試験や筆記課題(文献要約や外国語作文を含む)、後者では学期はじめに課題を提示しなるべく少人数で学習プロセスを観察しながらAI生成物からのコピーを防いで口頭発表やレポートによって評価することを例として挙げました。生成AIの利用の可否についてはシラバスに明記し、AIリテラシー教育を行う必要があると考えています。

生成AIの進歩は急速です。チャットポット・アリーナでの順位が頻繁に入れ替わる、AI戦国時代のような様相を呈しています。生成AIにできることもどんどん増えており、教員がそれを把握していないと、そもそも生成AI時代において、公平に評価できる課題の設定をすることはできません。そのため、教員が技術の進歩に目を向け続けていくことが大切ではないかと思っています。

[7] ハルシネーションについてや、学習・それ以外での生成AIの有効活用の方法・注意点などについても記述させた。