生成系 AI との相談を取り入れた数理データサイエンス AI 教育 プログラムの実践と検証

二瓶 裕之*1, 西牧 可織*1, 米田 龍大*1*1 北海道医療大学 情報センター

Practice and evaluation of MDASH incorporating tutoring powered by generative AI

Hiroyuki Nihei*1, Kaori Nishimaki*1, Ryuta Yoneta*1
*1 Health Sciences University of Hokkaido

内製化した知的学習支援システムに生成系 AI を API 経由で組み込むことで、数理データサイエンス AI 教育プログラムを構成する毎回の授業後に、学生が生成系 AI と相談できるようにした。目的の1つが、小さな疑問でも解決の糸口を随時提供することで、知る楽しさを伝えることである。もう1つが、日常的に生成系 AI に触れることに加えて、課題解決型学修においても生成系 AI への相談を組み込むことで、AI の誤りや欠点も繰り返し体験できるようにして、AI の回答への批判的観点を養うことである。

キーワード: 生成系 AI, API, 知的学習支援システム, 数理データサイエンス AI 教育プログラム

1. はじめに

教員には尋ねられないような小さな疑問であっても、 授業が進むにつれて学修の大きな障壁となってしまう ことがある。たとえ小さな疑問でも、随時、解決の糸 口を学生に提供することは重要である。しかし、その ために、教育支援者として常に人員を確保したうえで 安定した教育方針を提供することは難しい。

そこで、かねてより、知的学習支援システム (Intelligent Tutoring Systems、以下、ITS)(1)を活用したうえで、学生の抱える疑問に対応をしていくといった研究が盛んに行われてきている。ITS は、学生の問題や課題を段階的に解決する糸口を提供する機能を持ち、学生の状態に応じた適応的なフィードバックを提供するとされている。最近では、ITS にエージェントを導入した協働学修支援の研究(2)もされている。

このようなITSと教育の文脈で結びつくとして注目 を集めているのが Chat GPT などの生成系 AI である。 例えば、ITS に生成系 AI を組み込むことで、学生との 自然な対話を通じて教育的なフィードバックを提供したり、学生の相談や質問に対する解決の糸口を提供したりできるのではないかと期待されている。

一方で、大学教育の現場では、生成系 AI とどのように向き合うべきかが喫緊の課題ともなっている。例えば、東京大学では「生成系 AI(ChatGPT, BingAI, Bard, Midjourney, Stable Diffusion 等)について」とした意見がだされ、検索ではなく相談するシステムとして使うことを勧めること、信憑性に対する注意が必要であること、個人情報などの送信に留意することなどが挙げられている(3)。

このような中、北海道医療大学(以下、本学)では、DX 推進計画を策定して数理データサイエンス AI 教育プログラム(MDASH)を実施してきた(4)。特徴は、MDASHの中で、内製化してきた様々な教育支援システムを活用してきた点であり、ITS などの開発や活用も含まれている。例えば、自然言語処理に関わる機械学習システムを開発して、授業の振り返りコメントや講義ノートをクラスタリングしながら学生の学びを可

視化するなどしてきた。

そこで、「ITSにより生成系 AIへの相談を日常的に繰り返すことで、日々の疑問を解決する糸口をつかめるようになるのか」をメインのリサーチクエスチョンとして、内製化してきた機械学習システムに API 経由で Chat GPT を組み込んだ ITS を構築し、その日の授業内容について、学生が生成系 AI と相談できる仕組みを作った。直接生成系 AI を使わせるのではなく、APIを経由したのは、教員が相談内容の傾向を事前につかんで相談内容や、その適切性をチェックしたり、オプトアウトすることで大学外の組織が運用する生成系 AI に学生の学修データを学習させないためである。

今回の取り組みの目的の1つは、小さな疑問であっても、随時、解決の糸口を学生に提供することである。これにより、その日の疑問が授業のたびごとに解決されるようにし、また、それを知ることの楽しさへとつなげることである。もう1つが、生成系 AI からの回答に対する批判的な視点を養うことである。今回、日常的に繰り返し生成系 AI に触れることに加えて、課題解決型学修においても生成系 AI への相談を組み込んだ。これにより、AI の回答に含まれる誤りや AI の欠点も繰り返し体験できるようにし、批判的観点を持ちながら、生成系 AI を効果的に使えることを目指した。

2. 生成系 AI との相談を取り入れた授業概要

今回の取り組みの背景となった本学の DX 推進計画では、内製化した教育支援システムや LMS に、蓄積された学習ログの AI 解析機能を拡充することで、ITS を始めとして、小・中規模の機動性の高い AI を多種多面的に内製化してきた。一方で、多種多面的な AI が身近にあることからも、AI に振り回されることのないよう、全学的に、初年次に数理データサイエンス AI 教育プログラム(MDASH)リテラシーレベルを実施し、AI を活用できるスキルの醸成を図っている。

学生が生成系 AI と相談できる仕組みとしての ITS を活用した授業科目は、MDASH に含まれる「情報処理演習」である。「情報処理演習」は本学1年生に対する全学共通教育科目であり、全ての学部で必須科目となっている。履修者数は約 700 名である。「情報処理

演習」は、MDASH リテラシーレベルプラスに認定・ 選定されており(参考文献)、学生は、毎時間、BYOD として自身のパソコンを持参する。授業時間内には、 学生は、毎回、自身のパソコンを使って、その回のテ ーマに関した演習課題を google フォームから送信す ることとしている。

「情報処理演習」でITSを活用した理由は、生成系 AI との相談を始める前に、生成系 AI を含む AI 全般 の仕組み、革新性、そして、AI の欠点や活用するうえでの注意点などについて学ぶ機会を提供するためである。表 1 は「情報処理演習」の授業テーマである。数理データサイエンス AI をテーマとした授業回の1~3において、MDASHのスキルセットである「データ・AI の活用領域」、「データ・AI 利活用のための技術や最新動向」を扱うが、ここで、AI の仕組み、革新性、欠点などをグループワークも取り入れながら学ぶ。また、第3回目には、医療分野における AI 活用に関して調査した結果をレポートにまとめさせている。

表 1 「情報処理演習」の授業テーマ

授業回			授業テーマ
1	~	3	数理データサイエンスAI
4	~	6	プレゼンテーション
7	\sim	9	データ分析
10	\sim	12	データ処理、データ集計、データ分析
13	~	15	情報セキュリティとモラル

生成系 AI との相談を取り入れたのは5回以降の授業回とした。第5回と第6回では「プレゼンテーション」がテーマである。ここでは、第3回に作成したレポートを発表原稿とみなしたプレゼンテーションを作成させた。第7回から9回では「データ分析」がテーマである。ここでは、統計局の統計ダッシュボードの各種データを使って、健康や医療における課題や、今まで気づかなかったような広く社会に起きている現象や課題を見つけだす課題解決型学修(PBL)を実施した。

これらの授業回において、生成系 AI への相談を定期的に繰り返しながら、生成系 AI から得られた回答を学生に評価させて、その有用性、新規性、信頼性を判断させた。特に、PBL を実践する過程の中で生成系 AI への相談を組み込むことで、生成系 AI の回答も糸口にしながら、問題を解決する方策を立てられるよう

にした。これにより、生成系 AI からの回答を評価し、 最良の選択肢を選択しながら、批判的な観点を持つこ との価値も醸成されることを期待した。

ITS により生成系 AI への相談を実施した工程には

3. 生成系 AI への相談を実施した工程

4つのプロセスがあり、それぞれ、教員が学生の相談を受け付けるプロセス、相談内容を確認するプロセス、生成系 AI に相談を送信するプロセス、回答を受け取るプロセス、最後に、学生が回答を閲覧するプロセスである。これらのプロセスを、授業回ごとに繰り返すことで、学生の問題解決時の各ステップのレベルで支援する機能、学生の状態に応じた適応的にフィードバックする機能の2つをITSとして持たせるようにした。学生の相談を受け付けるプロセスでは、学生は、演習課題をgoogleフォームから送信するときに、あわせて「今日の授業で学んだこと」と「授業テーマに関して AI に相談してみたいこと」も記入するようにした(図1)。Googleフォームは、毎時間、同じフォームを利用して、相談の内容などは1つの google スプレッドシートに蓄積されるようにした。

回答を入力	
課題の提出	В
☆ ファ-	イルを追加
今日の授業 200文字)	僕テーマに関して生成系AIに相談したいことを記入してください(最大 ★
談するとき	具体的な問題や課題について他の人の意見やアドバイスを求めることです。相 には、自分の問題や課題を他の人に明確に説明し、それに対する回答や解決策 。相談することで、意思決定や問題解決のプロセスを助けられ、より広い視野

図1 学生の相談を受け付けるプロセス

相談内容を確認するプロセスでは、Google colaboratoryの開発環境で python を使って、google スプレッドシートから読み取った相談の内容を Doc2Vec で分散表現した後に、クラスタリングと異常 値検出をする。クラスタリングは非階層型と階層型の

2種類を行った。非階層型のクラスタリングでは、t-SNE により次元削減をしてから k-means 法によりクラスタリングし、クラスターの中央に位置した相談の内容を教員が閲覧できるようにした。階層型クラスタリングでは Ward 法によりクラスタリングして、クラスターごとに TF-IDF 値の高い単語を抽出することで、学生がどのような相談をしているのかの傾向をつかめるようにした。また、IsolationForest 関数を使った異常値検出により、適切性を欠いた相談が無いのかを教員が事前に確認できるチェック機能も設けた。

生成系 AI に相談を送信するプロセスでは、まず、Chat GPT へ送信するプロンプトを生成した。プロンプトには、学生ごとの「今日の授業で学んだこと」と「授業テーマに関して AI に相談してみたいこと」を変数で渡すとともに、共通のプロンプトとして、例えば、「相談したいことはないとした学生には、次回からは、学んだことをもとにして、何か質問をするように促してください」などの教員からの指示も含めた。

回答を受け取るプロセスでは、ChatCompletion. create コマンドにより、API 経由で、gpt-4 モデルにより回答を受け取った。API を利用してオプトアウトすることで、大学外の組織が運用する生成系 AI に学生の学修データを学習させないようにした。また、gpt-4 モデルの場合、回答にかかるトークンに対して従量制課金の費用が発生するために、回答は最大でも 500 文字以内とするように制限を設けた。回答の受け取りは学部ごとに一括して実施したが、1回のChatCompletion. create コマンドの実施のたびに、10 秒間のスリープを入れた。ChatCompletion. create からの応答がなかった場合には、さらに、10 秒間のスリープの後にリトライするようにした。

学生が回答を閲覧するプロセスでは、まず、Chat GPT から受け取った回答をデータベース化して、ITS から ODBC 経由で参照できるようにした。学生は ITS にログインすることで、Chat GPT からの回答を閲覧できるようにし、回答ごとにアンケートも実施した。また、回答をデータベース化することで、Chat GPT からの回答が他の授業科目のレポートに使われてしまうなどの剽窃行為への対応もできるようにした。

4. 実践結果に対する検証

4.1 検証の概要

ITS により生成系 AI への相談を日常的に繰り返すことで、日々の疑問を解決する糸口をつかめるようになるのか、また、批判的観点を持ちながら、生成系 AI をより効果的に使えるようになったのかを検証する。

4.2 相談の実施回数と閲覧回数の検証

生成系 AI への相談を日常的に繰り返せたのかを検証するために、図2に、生成系 AI への相談回数を示した。図2は、4学部の「情報処理演習」の授業回4回目から7回目までの相談回数を示しており、1週ごとに、ほぼ履修者数と同数となる心理科学部70回、リハビリテーション科学部150回、看護福祉学部150回、医療技術学部80回、あわせて500回の相談が繰り返されていることがわかる。なお、gpt-4モデルの従量制課金の費用は1週ごとに約20ドルであった。

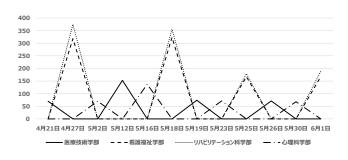


図2 相談回数

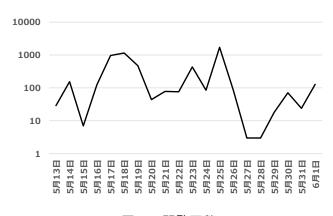


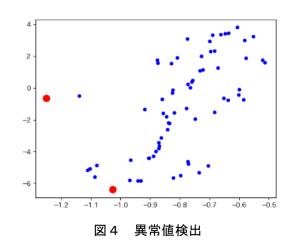
図3 閲覧回数

図 3 は、生成系 AI からの回答の閲覧回数である。ピークが現れるのは授業実施日である。例えば、2023/5/25 には、閲覧回数が約 1,700 回となっているが、これは、図 2 の 2023/5/18 の 200 件の相談を閲覧

している授業回に該当する。この結果から、相談開始 当初は、学生が何度も回答を繰り返し閲覧しているこ とがわかる。また、授業実施日以外にも、少ないなが ら、閲覧が記録されていることがわかる。なお、 2023/5/17 は、薬学部「文章指導」での利用であり、今 回の「情報処理演習」とは異なる授業科目での閲覧回 数となる。

4.3 チェック機能の検証

図4には、異常値検出をした例として、ある授業回における相談内容に対する結果を示した。グラフには、Doc2Vec により分散表現した相談内容を t-SNE により次元削減して、各々の相談内容の座標を直交座標にプロットしている。IsolationForest 関数により異常値検出をしたが、検出条件は contamination=0.02 とした。異常値とされた相談内容の座標を赤点によりマークしたが、検出された2件は「文の組み立て方や表現の死 k 多」(一部抜粋)と「なし」である。前者の文末部分は「表現の仕方」のタイプミスと考えられたため、生成系 AI ヘプロンプトを送信する前に修正している。



4.4 相談内容の検証

図5は、すべての相談内容をクラスタリングした結果である。図4と同じく、Doc2Vecにより分散表現した相談内容を t-SNE により2次元の直交座標にプロットしている。明確なクラスターは作られてはいないが、相談内容の概要を掴むために、k-means 法により8個のクラスターに分け、それぞれのクラスターの中央に位置した相談の内容を抽出した。抽出した相談内容(一部要約)は、下記の通りとなった。

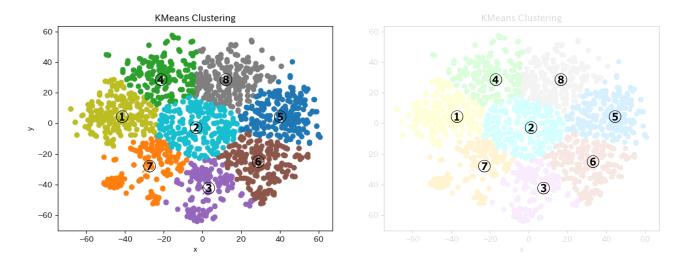


図5 すべての相談内容のクラスタリング(t-SNE と k-means): 右図はクラスター番号を強調表示

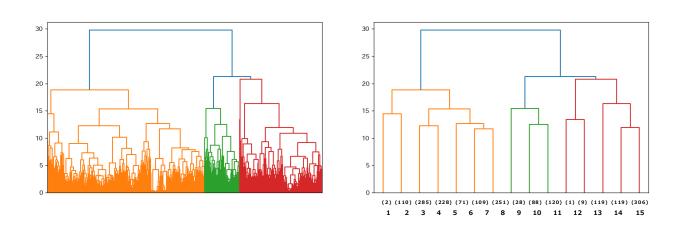


図 6 すべての相談内容のクラスタリング (Ward と dendrogram): 右図は 15 個にクラスタリング 表 2 クラスター1~15 の TF-IDF 値の高い単語 (%は全体に対するクラスターのサイズの比率)

			_								
		1	2	3	4	5	6	7	8	9	10
1	0 %	スライド	情報	不可能	人間	伝達	理解	視覚	ai	point	power
2	5 %	どう	いい	レポート	自分	よい	教え	ポイント	パワー	ai	ほしい
3	15 %	ポイント	パワー	教え	作り方	やすい	ください	スライド	コツ	プレゼンテーション	方法
4	12 %	方法	レポート	書き方	よい	スライド	文章	ai	ポイント	いい	プレゼンテーション
5	3 %	スライド	やすい	作る	文章	できる	どう	教え	レポート	知り	powerpoint
6	5 %	特に	なし	レポート	書く	スライド	どう	コツ	作る	知り	よい
7	13 %	ポイント	パワー	いい	やすい	スライド	プレゼンテーション	作る	どう	使う	よい
8	1 %	理由	なぜ	人口	北海道	東京	知り	増加	年代	看護	観光
9	4 %	人口	なぜ	日本	減少	高齢	ピラミッド	増加	くらい	どこ	また
10	6 %	なぜ	人口	増え	男性	減っ	サウジアラビア	多い	北海道	女性	ピラミッド
11	0 %	における	パワー	ポイント	リスト	有効	活用				
12	0 %	ai	医療	診断	できる	存在	早期	として	まし	分野	効率
13	6 %	ai	医療	活用	人間	思い	どう	できる	デメリット	れる	今後
14	6 %	人口	どう	変化	ai	日本	ピラミッド	医療	ください	なぜ	教え
15	16 %	教え	人口	ai	ください	医療	特に	日本	理由	方法	知り

- ① レポートのなかでの重要な点の書き方
- ② 良い PowerPoint の書き方
- ③ 見やすい PowerPoint の作り方
- ④ 東京に住んでいる人はどこで働いているのか
- ⑤ 文字数が決まっているときのレポートの書き方
- ⑥ AIの欠点とは
- ⑦ AI をうまく活用するにはどうするべきか
- ⑧ 生成系 AI をうまく活用するにはどうするべきか ①、②、③、⑤はレポートやプレゼンテーションの 作成方法に関して、⑥、⑦、⑧は医療分野における AI 活用に関して、④は「データ分析」をテーマとしたと きの課題解決型学修に関する質問であると考えられる。

図 6 は、Doc 2Vec により分散表現した相談内容をWard 法により階層型のクラスタリングをして、dendrogram ライブラリにより樹形図を示した。左は全ての質問内容のリーフを描画しているが、右が 14 個のクラスターに分けている。表 2 は、クラスター1~14の TF-IDF 値の高い単語である。

図5と図6の結果をまとめると、学生がどのような相談をしているのかの傾向をつかめる。例えば、プレゼンテーションの授業回では、スライドやレポートの作り方のコツや、授業のテーマであった医療分野における AI 活用の事例などに関する質問があることがわかる。また、統計局の統計ダッシュボードの各種データを使った PBL に関する質問も見受けられる。ただ、表2のクラスター6にあるように、一定数(全体の約5%程度)は、「質問がない」としている学生がいることもわかる。

4.5 新規性、信頼性、有効性の検証

次に、日々の疑問を解決する糸口をつかめるようになるのかを検証するために、生成系 AI からの各回答に対する学生からのアンケート結果を図7に示した。アンケートでは、生成系 AI からの各回答に対して、

- 1. 新規性がある(自分では気づかなかったような新しい視点が含まれていた)
- 2. 有用性がある(自分が抱える課題や問題を解決するのに役立った)
- 3. 信頼性がある(正確で信頼できると感じた)

の3点について、それぞれ、強くそう思う、そう思う、 そう思わない、強くそう思わないの4件法の回答を求 めた。この結果、全体的に、各回答に対して、有用性 が最も高く評価されており、それと比較すると、信頼 性や新規性は低くなる。



図7 生成系 AI からの各回答にアンケート結果

有用性よりも信頼性が低くなる点は、批判的観点を持ちながら、生成系 AI をより効果的に使えるようになるきっかけになると考えた。そこで、PBL の過程で学生が生成系 AI への相談したことへの回答に対して、「Chat GPT の回答が不十分であった点を書いてください」、「Chat GPT の回答が正しいかを確認するために追加で探した情報があれば書いてください」の 2 点について、自由回答のオープンエンド形式でのアンケートを実施した。

なお、オープンエンド形式であることから、アンケート結果の集計においては、ITS によるクラスタリング、ならびに、gpt-4 モデルによるクラスターごとの要約を実施した。

全ての学部で、「Chat GPT の回答が不十分であった 点を書いてください」に対して回答があったのは 113 件であった。表 3 には、アンケート結果をクラスタリングして、クラスターの要約文を件数と併せて示した。同様に、表 4 には、「Chat GPT の回答が正しいかを確認するために追加で探した情報があれば書いてください」に対するアンケート結果を示した。この質問に対する回答があったのは 65 件となり、全履修者数の 2 割程度の回答であった。

表3 「Chat GPT の回答が不十分であった点を書いてください」に対するアンケート結果

クラスターごとの意見の要約	件数
特に不十分な点はなかった・特になし	33
回答が短く・具体例や詳細が不足していた	14
回答が求めていたものと違った・欲しい回答が得られなかった	9
調査結果や回答の信憑性・正確性に疑問があった	8
既知の情報やありきたりな回答が多かった	7
質問と関連性が薄い回答があった	6
言葉の意味が分からない表現やわかりにくい部分があった	5
質問に対する回答が理解できたが、追加情報が欲しかった	4
質問や回答の内容に誤り・誤差があった	4
質問に対する回答がなかった・回答が見つからなかった	3
その他の意見	20

表 4 「Chat GPT の回答が正しいかを確認するため に追加で探した情報があれば書いてください」に対す るアンケート結果

	m. su.
クラスターごとの意見の要約	件数
人口や出生率に関する情報	9
歴史・政策に関する情報	9
情報の正確性を確認	7
経済・労働に関する情報	6
移民・移住に関する情報	5
医療・看護に関する情報	4
食料・農業に関する情報	3
その他	22

表3と結果から、まず、Chat GPTの回答が不十分ではなかったとした意見が3割程度あったが、7割程度は、具体性や信頼性に対する指摘が挙げられている。表4の結果からは、PBLのテーマとなった情報があがっており、hat GPTの回答が正しいかを確認するために追加で探した情報としては、適切な事項であった。しかし、全体の2割程度の回答であったことは、今後の大きな検討課題となる。

5. むすび

「ITS により生成系 AI への相談を日常的に繰り返すことで、日々の疑問を解決する糸口をつかめるようになるのか」をメインのリサーチクエスチョンとして、内製化した知的学習支援システムに生成系 AI を API 経由で組み込むことで、MDASH を構成する毎回の授業後に、学生が生成系 AI と相談できるようにした。

検証をはじめてから一か月程度の期間ではあるが、

相談回数は週に 500 回程度、合計で 3,000 回を超え、閲覧回数も週に数千回となるなど、安定的に、相談の受付がされていた。相談内容は、授業内容に沿った相談が多く、プレゼンテーションの授業回では、スライドやレポートの作り方のコツや、授業のテーマであった医療分野における AI 活用の事例などに関する質問が多かった。ただ、一定数は、「質問がない」としている学生もいた。

日々の疑問を解決する糸口をつかめるようになった のかを検証した結果としても、AIからの回答に対して、 有用性が最も高く評価されていた。しかし、有用性と 比較すると、信頼性や新規性は低くなった。

有用性よりも信頼性が低くなる点は、批判的観点を持ちながら、生成系 AI をより効果的に使えるようになるきっかけになると考える。そこで、Chat GPT の回答が不十分であった点や Chat GPT の回答の正しさを確認するために追加で探した情報がどの程度あるのかも検証した。結果は、2割程度の学生が、Chat GPTの回答に対して、適切な情報を自身で追加して探していることがわかった。しかし、まだ、このような学修行動の広がりはまだ十分ではないと考える。

一方で、著者らは、他の授業科目において、異なる 授業形態で同様の取り組みをしており、ここでは、批 判的観点を持つことの価値が醸成されつつあるとの結 果も出始めている。今後は、実践と検証を重ねて、批 判的観点を持ちながら生成系 AI を効果的に使えるよ うなスキルが修得されることを目指したい。

謝辞

本研究は JSPS 科研費 22H01051 の助成を受けた ものです。

参考文献

- (1) 林 勇吾: "私のブックマーク「知的学習支援システム (Inteligent Tutoring Systems)」",人工知能, Vol. 33, No. 4, pp. 527-530 (2018)
- (2) 吉川 厚, 高橋 聡: "エージェント技術の教育応用", 教育システム情報学会誌, Vol. 35, No. 1, pp. 5-12 (2018)
- (3) 東京大学 生成系 AI(ChatGPT, BingAI, Bard,

Midjourney, Stable Diffusion 等)について, https://utelecon.adm.u-tokyo.ac.jp/docs/20230403generative-ai (2023 年 6 月 9 日確認)

(4) 二瓶 裕之, 西牧 可織: "医療系大学における数理データサイエンス AI 教育 (リテラシーレベルプラス) の実践と検証", 教育システム情報学会誌, Vol. 40, No. 2, pp. 176-178 (2023)