

科目特化型データサイエンス教育に向けた RAG ベース Q&A システムの構築

竹内光悦*1・山本幸太郎*2・町田 保*3・末永勝征*4

Email: takeuchi-akinobu@jissen.ac.jp

- *1: 実践女子大学人間社会学部
- *2: 株式会社想隆社
- *3: リトルスタジオインク株式会社
- *4: 鹿児島純心女子短期大学

◎Key Words 生成 AI, データサイエンス教育, 統計教育

1. はじめに

初等中等教育における統計・データサイエンス教育が拡充され、2025 年度からは、その学習指導要領に基づく教育を受けてきた学生が大学に入学している。また、高等教育においても、理系・文系を問わずデータサイエンス教育を受ける機会が増えており、多くの大学で導入に向けた取り組みが進められている。一方で、一般に統計・データサイエンスを苦手とする学生が多いとされる文系学生の中には、これらを「数学」の一分野と捉え、強い苦手意識を持つ者も見られる。このような課題への対応として、国内外でさまざまな研究が行われている。

著者らの先行研究においても、竹内 (2020) では、文系女子大における社会調査および調査データ分析を中心とした体系的な学習カリキュラムの実装を報告している。また、竹内 (2022b) では文系女子大学でのデータサイエンス教育の導入に伴う課題を取り上げた。さらに、竹内 (2022a) では AI 時代に求められる統計教育のあり方を議論し、竹内 (2024) では生成 AI の活用事例を報告している。竹内 (2025a) では、統計教育に限らず、生成 AI を活用した授業設計が今後一層重要になることを指摘している。さらに竹内 (2005b) 文系学生を対象とした LMS (学習管理システム) のデータを活用した学習支援についても紹介している。また、日本統計学会統計教育分科会・JDSSP 高等学校データサイエンス教育研究会 (2025) では、ChatGPT を活用したデータ分析教育の導入事例を多角的に紹介しており、今後もこうした動きが加速していくと考えられる。実際、統計・データサイエンス教育に限らず、授業法・教材開発、質問応答、課題や成績の処理などに生成 AI を活用する研究報告も見られる (第 22 回統計・データサイエンス教育の方法論ワークショップ事務局、2025)。さらに、数学に特化した生成 AI ツール (Mathful、2025) の登場により、これまでにない授業展開の可能性も広がっている。

このような状況を踏まえると、生成 AI の活用により、ワンツーマン指導のような個別対応が可能となるだけでなく、文系学部にも多い大人数授業においても、24 時間 365 日対応の質問応答や中間課題の早期評価など、新しい教育のかたちが期待される。ただし、こうした取り組みに関する本格的な研究成果は、現時点では多く報告されていない。

以上を背景として、本研究では、特に科目特化型のデー

タサイエンス教育を支援するため、RAG (Retrieval-Augmented Generation) 技術を活用した Q&A システムの構築を試みる。

2. 科目特化型データサイエンス教育に向けた RAG ベース Q&A システム

本章では科目特化型データサイエンス教育に向けた RAG ベース Q&A システムの概要を述べる。

2.1 生成 AI を活用した Q&A システムへの期待

上述したように、データサイエンス教育は学校教育において徐々に普及しつつあるが、これまでの学習課程では同様の内容が組み込まれてこなかった。そのため、指導する教員自身がこうした分野を履修してこなかった場合、自信を持って指導できずに戸惑う声も、関係者へのヒアリングなどでしばしば耳にする。

また、関連する資料は多く存在するものの、一般向けのものや専門的な書籍との間には乖離があり、適切な教科書や参考書を見つけられず、初学者が戸惑う場面も多い。これらは、統計・データサイエンス教育を専門とする教員が担当すれば解決可能な問題ではあるが、近年の急激な関連学部・科目の新設に伴い、学外の非常勤講師や隣接分野の教員が担当している大学も多く、対応の質に差が出ているのが実情である。

一方、現在ではインターネット上に多くの情報があふれており、ウェブ検索やオンラインジャーナル、YouTube などの動画配信を通じて、丁寧な解説を得ることも可能である。加えて、X (旧 Twitter) などを用いれば、質問に対して多数のユーザーからさまざまな回答が寄せられる。しかしながら、このような情報源が豊富な一方で、情報が過剰で取捨選択に困り、情報の信頼性に不安が残るといった問題もある。

学習者の側では、高校までにある程度学習してきた学生も多いが、特に文系の学生には、高校 1 年次の「数学 IA」以降、数学から離れてしまっている者も少なくない。確率分布や統計的仮説検定など、統計・データサイエンスに関連する内容は「数学 B」に含まれているが、それを選択していない学生も一定数おり、非公開調査でもその実態が確認されている。

そのような学生にとって、教員への質問がしづらいつ感じ、質問できる環境が整っていても「簡単すぎる質問で

はないか」と遠慮してしまう場合がある。また、多様な生活スタイルを持つ学生にとっては、24 時間 365 日、すぐに質問できる環境があれば、学習の停滞を防ぎ、課題にも継続的に取り組むことが可能となる。

このような課題に対する一つの解決策として、生成 AI を活用した質問応答があげられる。しかし、現時点において生成 AI が常に正確な回答を出力するとは限らず、その真偽を確認するために余計な時間を要することもある。

さらに、統計・データサイエンスは数学だけでなく、医学、農学、生物学、工学、マーケティングなど幅広い分野で活用されているため、用語の「ゆらぎ」も存在する。たとえば、「標本の大きさ」は「標本数」や「サンプルサイズ」など複数の言い方があり、受講している授業での定義とネット上の情報との違いが、混乱を生む原因となっている。

こうした課題への対処法の一つとして、受講している科目の内容に特化した生成 AI システムを構築することが考えられる。本研究では、これらの背景を踏まえ、RAG を用いた Q&A システムの構築を目指す。

2.2 科目特化型データサイエンス教育に向けた RAG ベース Q&A システム

一般的な RAG を用いたシステム（図 1）では、AI に参照させるためのデータを指定することで、その授業内容に即した信頼性の高い回答を得られる可能性が高まる。

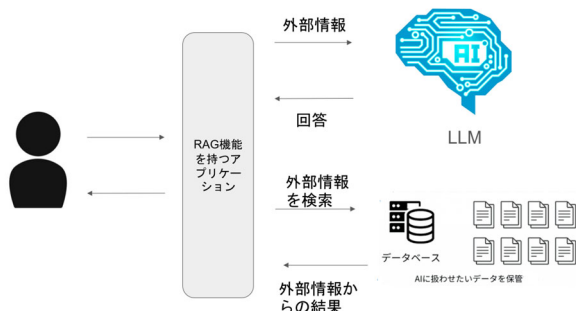


図 1. 一般的な RAG を用いたシステム
(引用：株式会社想隆社作成資料)

本システムによって、当初あげた課題は解決可能である。しかし一方で、学生の疑問点に関する情報が取得できず、授業改善に活用できなくなるという別の課題も生じる。

そこで本研究では、RAG ベースの Q&A システムに加えて、学生から寄せられた質問内容や回答までのやり取りを、個人が特定されない形で収集・蓄積する仕組みを組み込む。この情報をもとに、どのような質問が寄せられ、どのような対話を経て理解に至ったかを分析し、今後の授業改善に活かすことを目的とする。現在、図 2 に示すシステムの開発を進めている。

このシステムの活用により、学生が疑問を抱えたまま学習を進めることなく、適切に理解を深めながら学びを進めることが可能となる。また、時間や場所に制限されない質問環境により、理解度や定着度の向上も期待される。

さらに、従来のような授業後の質問対応だけでなく、学生が何気なく抱く小さな疑問の蓄積も可視化できるようになり、これまで見えなかった授業設計上の課題を新た

に発見する手がかりともなり得る。

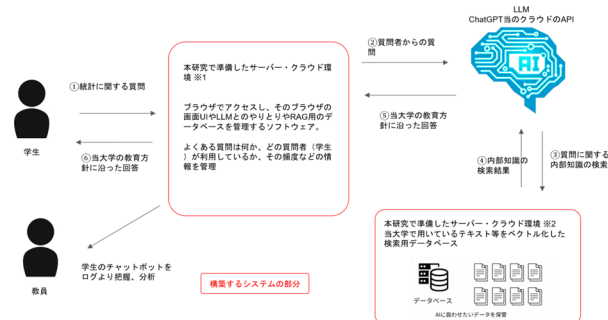


図 2. 統計授業のサポートチャットボットシステム
(引用：株式会社想隆社作成資料)

3. まとめ

今回の発表では、著者らが現在開発を進めている、科目特化型データサイエンス教育を支援する RAG ベースの Q&A システムの概要を紹介した。システムは現在も開発段階にあるため、実際の検証データについては、今後の発表機会にて報告する予定である。

なお、生成 AI に関連する分野は急速に発展しているため、その動向を適宜踏まえながらシステム設計や運用方針を柔軟に調整していくことが求められる。本研究においても、こうした点に留意しながら今後の開発を進めていく予定である。

参考文献

- (1) 第 22 回統計・データサイエンス教育の方法論ワークショップ事務局 (2025) 第 22 回 統計・データサイエンス教育の方法論ワークショップ～デジタル人材育成を視野に入れた統計・データサイエンス教育を考える―新学習指導要領を見据えた生成 AI や探究学習を活用した統計・データサイエンス教育における文理横断・文理融合教育の展開一、<https://estat.sci.kagoshima-u.ac.jp/SESJSS/edu2025.html> (最終確認日：2025/05/20)。
- (2) 竹内光悦 (2020) 社会調査・分析を中心とした女性データサイエンス教育の展開、日本ソーシャルデータサイエンス学会論文誌、4(1)、13-18。
- (3) 竹内光悦 (2022a) AI 時代に必要な統計教育―学校教育から見る AI 時代に必要な統計教育の 3 要素の学び―、コンピュータ&エディケーション、52、12-17。
- (4) 竹内光悦 (2022b) 文系女子大学におけるデータサイエンス教育の導入と実施の課題、実践女子大学人間社会学部紀要、18、111-117。
- (5) 竹内光悦 (2024) データサイエンス教育における生成 AI の利活用、2024 年度数学教育学会秋季例会、数学教育学会、89-91。
- (6) 竹内光悦 (2025a) 公開講座 AI・データサイエンス教育とこれからの高等教育、人と教育、目白大学高等教育研究所、47-64。
- (7) 竹内光悦 (2025b) 文系学生を対象とした統計学習支援における質問・アクセスデータの活用、日本分類学会第 44 回大会、19-20。
- (8) 日本統計学会統計教育分科会・JDSSP 高等学校データサイエンス教育研究会 (2025) データ分析の基礎 生成 AI ではじめる統計・データサイエンス、東京図書。
- (9) Mathful (2025) MathGPT、<https://mathful.com/ja/mathgpt> (最終確認日：2025/05/20)。