

# 大規模私立総合大学における 生成 AI を用いたデータサイエンス・AI 教育

岡部格明\*1・木元篤郎\*1・宿久洋\*1  
Email: mokab.0328@gmail.com

\*1: 同志社大学

◎Key Words

数理・データサイエンス・AI 教育, 大規模授業, 電子教科書

## 1. はじめに

情報技術の発達により、大量のデータを収集、解析することが可能となったことにより、AI やデータサイエンスに対する社会の期待が高まっている。AI・データサイエンスの利活用も広がりつつあるが、依然として、数理・データサイエンス・AI を身につけた人材が不足していることが指摘されている。2019 年に政府は、AI 戦略 2019<sup>(9)</sup> をまとめた。この中では、AI、数理、データサイエンスの具体的な教育目標として、「文理を問わず、全ての大学・高専生（約 50 万人卒/年）が、過程にて初級レベルの数理・データサイエンス・AI を習得することが、挙げられている。同志社大学では、これを受けた取り組みとして、2022 年度より、同志社データサイエンス・AI 教育プログラム（DDASH プログラム）として、数理・データサイエンス教育強化拠点コンソーシアムの策定したプログラムのリテラシーレベルに準拠したプログラムを全学で実施している<sup>(10)</sup>。

一方で、近年、生成 AI がより普及し、教育分野における生成 AI の活用も課題となっている。文部科学省では、大学・高専における生成 AI の教学面の取扱いについての声明を発表している<sup>(13)</sup>。このほかにも、東京大学や大阪大学なども、生成 AI の講義や学習での活用に関する声明を発表している<sup>(8)(9)</sup>。このように、生成 AI を利活用した教育自体が大きな関心ごととなっている。その中で、大規模言語モデルを用いた生成 AI では、事実ではない情報を事実のように提示してしまう、ハルシネーション（幻覚）を引き起こすことがある<sup>(2)</sup>。さらに、一般向けに公開されている生成 AI のサービスでは、セキュリティの観点から入力する情報に対して注意が必要な場合がある。そのため、教育的な用途での生成 AI システム の利用においては、注意深く検討する必要がある。

このような状況の中で、同志社大学では、NTT 西日本、NTTEDX と共同で、生成 AI を活用したデータサイエンス・AI 教育の実証事業を行っている<sup>(11)</sup>。この実証事業では、次の事項に焦点を当てた生成 AI 利活用の検証を行っている。具体的には、学生への学習支援の仕組み、教職員の教育支援の仕組み、生成 AI の回答を電子教科書・電子教材に限定した利用検証、生成 AI の仕組みを学習向けにチューニングするノウハウの展開の 4 つの事項に焦点を当てている。この実証事業の中で、生成 AI を用いた学習支援システムである、Doshisha AI Buddy (DAIB)<sup>(12)</sup> を NTT 西日本と共同で開発した。本報告では、この DAIB を用いたデータサイエンス・AI 教育の取り組みについて報告する。

## 2. DAIB

DAIB は、Microsoft Azure サービスを用いて構成されている、生成 AI を用いた AI アシスタントである。DAIB は、Microsoft Teams 上のチャットをインターフェイスとして用いたシステムである。DAIB のインターフェイス画面を図 1 に示す。利用者は、Microsoft Teams のチャットに DAIB を追加し、利用を開始することを想定している。

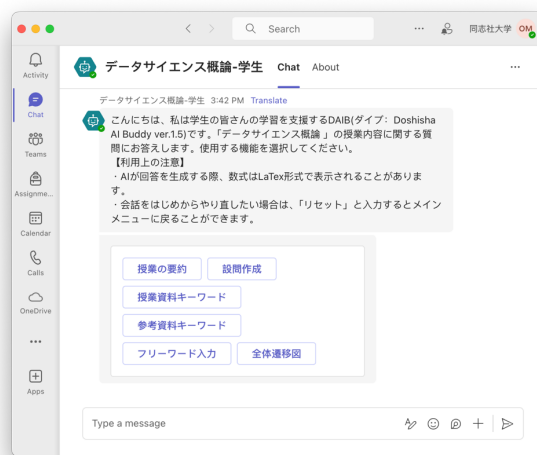


図 1 DAIB のインターフェイス画面

DAIB では、情報源として用いる情報をホワイトリストとして教員側が制限をしているため、信頼できる情報を利用した回答を行うことができることが特徴である。信頼できる情報源としては、DAIB を用いる講義のために作成された講義資料や教科書のほか、数理・データサイエンス・AI 教育強化拠点コンソーシアムが公開している講義資料やモデルカリキュラムを用いている。コンソーシアムが公開している講義資料については、PDF 公開されている教材を用いている。情報源である教科書の利用に関して、DAIB は、NTTEDX 社が提供する電子教科書配信サービス「EDX UniText」<sup>(4)</sup>と連携することで、教育・学習向けのセキュアな環境で利用できることがもう一つの特徴として挙げられる。

次に DAIB の機能について説明する。DAIB の全体遷移図を図 2 に示す。学生向けに利用可能な DAIB の学習支援機能には次のものがある。1 つ目は、授業内容の要約である。これは、簡易要約と詳細要約の 2 つの要約を利用可能であり、利用者に想定される要求としては、初学者や

大枠を掴みたいことを想定した場合の簡易要約と、より詳細な内容を知りたい場合の詳細要約がある。詳細要約では、数式を用いたより詳細な出力を得ることができる。2つ目は、授業内容に関する設問作成である。これは、授業内容をもとにした、選択肢式問題や穴埋め式問題、記述式問題を生成する機能である。生成と同時に答えが出力され、利用者は、生成された問題を利用して学習を進めることができる。3つ目は、授業内容に関するキーワード解説である。これは、授業内容に関するキーワードを解説する機能であり、利用者は、授業内容に関連するキーワードのリストの中から知りたいキーワードを選択することで、そのキーワードに関する解説を得ることができる。4つ目は、フリーワード入力による質問対応である。これは、利用者が自由に質問を入力することで、その質問に対する回答を生成する機能である。このときにも、回答生成には信頼できる情報源を用いているため、講義の内容や教科書の内容に則した回答を得ることができる。

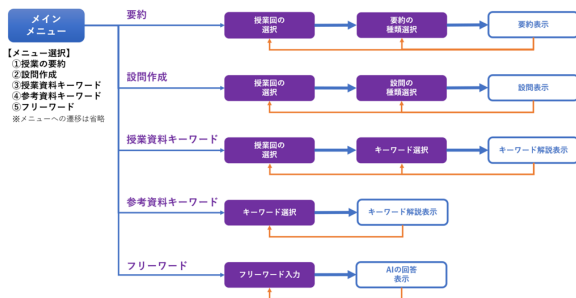


図 2 DAIB の機能遷移図

内部の生成 AI には、GPT-4 Turbo<sup>1)</sup> がベースとなっている Azure Open AI が用いられている。利用者は Microsoft Teams を介して Azure OpenAI にアクセスし、Copilot Studio<sup>3)</sup> を利用して過去の応答を用いた会話を行うことができる。そのため、各機能で得られた回答に対して、より詳しく知りたい場合や、得られた出力に対して質問がある場合に、継続して会話を行うことができる。

### 3. 実証事業対象科目

この実証事業では、DDASH プログラムを構成する科目である「データサイエンス概論」と「データサイエンス基礎」を対象としている。

DDASH プログラムの全体について、図 3 に示す。DDASH プログラムは DDASH-L (リテラシーレベル)、DDASH-A (応用基礎レベル)、DDASH 副専攻、DDASH-E (エキスパートレベル) の 4 段階で構成されている。DDASH-L, DDASH-A, DDASH 副専攻は学部レベルのプログラムであり、DDASH-E は大学院レベルのプログラムである。リテラシーレベルに対応する DDASH-L は、実社会で目にするデータを適切に読み解き、使い方を判断できる水準の数理・データサイエンス・AI に関する能力取得を目的としたプログラムである<sup>10)</sup>。この教育プログラムは、必修科目である「データサイエンス概論」(2 単位) の単位取得に加えて、プログラム選択科目として、全学教養科目として開講されているいくつかの科目の単位を取得することにより修了することができる。

応用基礎レベルに対応する DDASH-A は、数理・デー

タサイエンス・AI 教育 (リテラシーレベル) の教育を補完的・発展的に学びデータから意味を抽出し、現場にフィードバックする能力や、AI を活用し課題解決につなげる機能力を修得し、自らの専門分野に数理・データサイエンス・AI を応用するための大局的な視点獲得を目的としたプログラムである<sup>10)</sup>。この教育プログラムは DDASH-L の修了要件に加えて、必修科目として「データサイエンス基礎」、「データエンジニアリング基礎」、「AI 基礎」(各 2 単位) の 3 科目の単位を取得することによって、修了することができる。

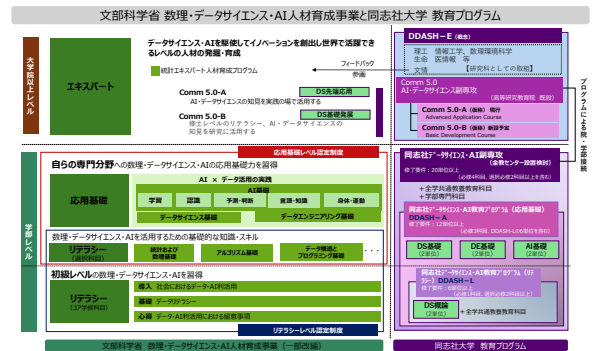


図 3 DDASH プログラムの構成図

実証事業対象の 2 科目の「データサイエンス概論」と「データサイエンス基礎」は、それぞれ DDASH-L と DDASH-A の必修科目である。いずれの科目も同志社大学全学の学生が受講することができる科目であり、学期期間中にオンデマンドで開講される。

データサイエンス概論は、数理・データサイエンス教育強化拠点コンソーシアムにおいて策定された、リテラシーレベルに含まれる、「導入」、「基礎」、「心得」の項目に準拠した内容となっている<sup>6)</sup>。この科目は、「ガイダンスと導入」、「社会におけるデータ・AI の利活用」、「データリテラシー」、「データ・AI 利活用における留意事項」の 4 つの項目から構成される。教科書として、実教出版の「事例でわかる統計 教養のための統計入門」を用いている。この科目の中では、データサイエンスの方法に関わる箇所を扱っている「データリテラシー」の領域を生成 AI による学習支援を行う対象箇所としている。

データサイエンス基礎は、数理・データサイエンス教育強化拠点コンソーシアムにおいて策定された、応用基礎レベルに準拠した内容となっている<sup>7)</sup>。この科目はデータ分析の演習をすることに重点が置かれており、R 言語を用いたデータ分析の演習を行う。この科目は、「分析設計」、「データ観察」、「データ分析」、「データの可視化」の 4 つから構成されている。教科書としては、データサイエンス概論で用いられているものに加えて、実教出版の「事例でわかる統計 経済・経営計のための統計入門」を用いている。これに加えて、R の演習に関する参考書として、「多次元データ解析法 (R で学ぶデータサイエンス 2)」を用いている。この科目の中では、講義の中で扱う全体的内容を生成 AI による学習支援を行う対象箇所としている。

#### 4. 受講者の利用状況

2024 年度春学期におけるそれぞれの講義の受講者は、データサイエンス概論が 2416 人、データサイエンス基礎が 124 人である。2024 年 6 月 24 日現在で、この実証事業の対象科目としている講義の受講者のうち、データサイエンス概論では、1565 人が利用しており、これは、講義登録者 2416 人のうち 64.8%が利用していることになる。データサイエンス基礎では、78 人が利用しており、これは、講義登録者 124 人のうち 62.9%が利用していることになる。

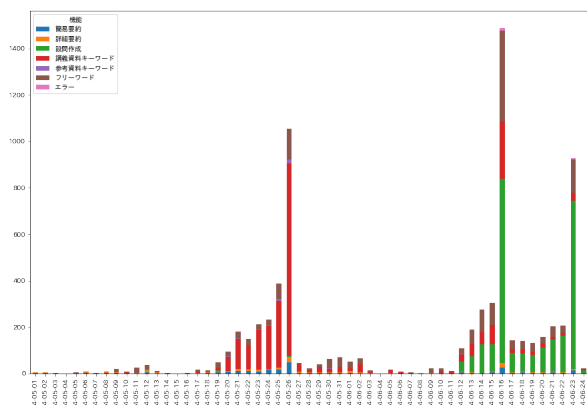


図 4 データサイエンス概論における DAIB の利用状況

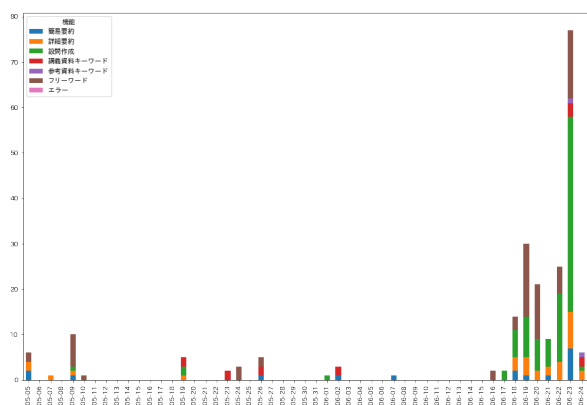


図 5 データサイエンス基礎における DAIB の利用状況

2024 年 5 月 1 日から 2024 年 6 月 24 日までで利用された機能の内訳を示した利用状況について、データサイエンス概論のものを図 4 に、データサイエンス基礎のものを図 5 にそれぞれ示す。いずれの科目でも、リリース時点からの利用はあまり多くなかった。データサイエンス概論では 5 月 20 日の週に配信された講義において、DAIB の講義資料キーワードを用いて回答を行うことを求めた課題を出題したところ、利用が促進されたことがわかる。多くの受講生がこの課題に取り組む際に初めて DAIB を利用したと考えられる。5 月 20 日の週に配信された講義で出題された課題の締め切りが 5 月 26 日であるが、以降もキーワード解説やフリーワードを中心に若干の利用が見られることから、ある程度継続的に利用をしている受講生の存在が考えられる。データサイエンス基礎では、6 月 17 日の週に配信された講義において、DAIB の設問作成を用いて回答を行うことを求めた課題を出題した。ここで多くの受講者が設問作成・フリーワードを中心とし

た機能の利用をしたことが考えられる。また、2 つの科目において特に日曜日に利用数が増加していることがわかる。これは、2 つの科目の課題の締め切りを毎週日曜日の 23:59 としているためであると考えられる。

データサイエンス概論で特に利用が多かったキーワード解説で調べられることが多かったキーワードを表 1 に示す。第 6 回講義において、課題として出題したところ、最も解説されたキーワードは「ジニ係数」であった。これは、受講者にとって最も馴染みがないキーワードであるためと推察される。第 7 回、第 8 回については課題としての出題は行っていない。第 7 回で多く解説された「散らばりの指標」については、講義内で扱われた「分散」や「標準偏差」などの散らばりの指標に興味を持った受講生の存在が推察される。第 8 回については、利用の総量が少なかった。これは、扱う内容自体が、高校数学にも含まれる範囲である 2 つの量的変数の関係性に関する範囲であり、受講生にとっても馴染み深かったからであるためと考えられる。第 9 回については、設問作成に関する課題を出題しているが、キーワードとして「クラメールの連関係数」が調べられている。第 9 回の内容は 2 つの質的変数の関係性に関するものであったことから、受講生にとっては、新たな概念の登場が多く、第 8 回に比べて難易度が上がったことが原因として考えられる。

表 1 データサイエンス概論においてよく解説された授業資料キーワード

第 5 回	第 6 回	第 7 回
5 月 13 日公開	5 月 20 日公開	5 月 27 日公開
社会調査	ジニ係数	散らばりの指標
データリテラシー	平均	平均
前向き研究	ローレンツ曲線	標準偏差
母集団と標本	中央値	分散
標本抽出法	代表値	標準得点
第 8 回	第 9 回	第 10 回
6 月 3 日公開	6 月 10 日公開	6 月 17 日公開
共分散	クラメールの連関係数	箱ひげ図
偏相関係数	分割表	データ視覚化
相関係数	オッズ比	ヒストグラム
散布図	精度	モザイクプロット
量的変数	データリテラシー	散布図行列

そのほかにも、データサイエンス概論・データサイエンス基礎の DAIB において、「フリーワード」機能で質問されていた項目には、「偏差値の標準偏差が 10 になることをより詳しく教えてほしい」、「分割表とクロス集計表と相関表の違いを教えてほしい」、「畳み込みニューラルネットワークの原理を教えてほしい」、「偽陽性とは何か」などの質問が挙げられていた。このほかにも、R のコードに関するエラーの質問や自身が書いた R コードがあっているかを確認するような質問が見られた。



## 5. 今後の方向性

今後のDAIB開発・活用の方向性として、次の3点が挙げられる。

1つ目は、ホワイトリストとして参照する教材の拡充である。現状として、数理・データサイエンス教育強化拠点コンソーシアムに公開されている教材にとどまっているが、インターネット上に公開されており、利用可能なデータサイエンス教材はこのほかにも挙げられる。限られた資料の中からの出力では、キーワード解説において解説される内容の乏しさや、設問作成における問題の多様性の乏しさが問題となる。DAIBによる質問回答を豊かにしていくために、現状のものからいくつか加えていくことが今後の課題として考えられる。

2つ目は、出力内容の改善である。現状のシステムにおいては、利用者の質問から参照している教材へのリンクに関してはある程度信頼できるものである。しかし、参照内容から生成される解説については、信頼性の乏しいものが混じっていることがある。そのほかにも、設問作成においては、資料中の例に過適合した内容が出力されることもある。そのため、講義の課題として出題する際にも、出力内容には注意をすること、出力内容に意見がある場合には回答の中でそれを述べることを受講生に求めている。実際、出力内容に関して挙げられた意見として、「問題として問題文だけからは回答不能である」という意見が挙げられていた。そのほかには、利用者の理解が不十分であることに起因して、実際には問題ない出力に対しても、DAIBの出力に問題があるのではないかという指摘も挙げられていた。出力内容の改善のために、講義期間終了後の受講生のフィードバックを加味した改良を検討していく予定である。

3つ目は、教員向け機能の開発である。同志社大学が取り組んでいる実証事業の中で現在取り組んでいる学生向けのDAIBの他に教職員の教育支援の仕組みについても焦点を当てている。生成AIを活用する方向として、講義で用いる教材の作成や受講生の理解度評価のための設問作成が考えられる。これを中心としたよりよい学習のための教員向けシステムに関しても検討を進めていく予定である。

## 6. おわりに

本稿では、大規模私立総合大学である同志社大学における生成AIによる学習支援システムを用いたデータサイエンス・AI教育の取り組みについて説明した。より詳細なDAIBの機能や出力、利用状況の分析結果や春学期講義終了後の受講生へのアンケート結果などについては当日報告する。

## 参考文献

- (1) Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Al-tenschmidt, J., Altman, S., Anadkat, S. et al.(2023). Gpt-4 technical report, *arXiv preprint arXiv:2303.08774*.
- (2) Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., and Fung, P.(2023). Survey of hallucination in natural language generation, *ACM Computing Surveys*, **55**(12), 1–38.
- (3) Microsoft. 『 Microsoft Copilot Studio 』 .

- <https://www.microsoft.com/ja-jp/microsoft-copilot/microsoft-copilot-studio>.
- (4) NTT EDX. 『NTT EDX 事業・サービス紹介』 . <https://www.nttedx.co.jp/service>.
- (5) 首相官邸政策会議 (2019). 『AI 戦略 2019』 . [https://www.kantei.go.jp/jp/singi/ai\\_senryaku/pdf/aistrategy2019\\_fu\\_sanko.pdf](https://www.kantei.go.jp/jp/singi/ai_senryaku/pdf/aistrategy2019_fu_sanko.pdf).
- (6) 数理・データサイエンス教育強化拠点コンソーシアム (2020). 『数理・データサイエンス・AI(リテラシーレベル)モデルカリキュラム』 . [http://www.mi.u-tokyo.ac.jp/consortium/pdf/model\\_literacy.pdf](http://www.mi.u-tokyo.ac.jp/consortium/pdf/model_literacy.pdf).
- (7) 数理・データサイエンス教育強化拠点コンソーシアム (2021). 『数理・データサイエンス・AI(応用基礎レベル)モデルカリキュラム』 . [http://www.mi.u-tokyo.ac.jp/consortium/pdf/model\\_ouyoukiso.pdf](http://www.mi.u-tokyo.ac.jp/consortium/pdf/model_ouyoukiso.pdf).
- (8) 大阪大学. 『生成 AI 教育ガイド』 . [https://www.tlsc.osaka-u.ac.jp/project/generative\\_ai/](https://www.tlsc.osaka-u.ac.jp/project/generative_ai/).
- (9) 東京大学. 『生成 AI(ChatGPT 等)関連情報』 . <https://utelecon.adm.u-tokyo.ac.jp/online/topics/generative-ai>.
- (10) 同志社大学 (2023). 『「同志社データサイエンス・AI 教育プログラム」 (DDASH) 』 . <https://cgle.doshisha.ac.jp/ddash/overview.html>.
- (11) 同志社大学 (2023). 『同志社大学, NTT 西日本, NTTEDX 教育・学習活動への生成 AI 活用 実証事業スタート! ~教育・学習向け生成 AI を活用した新たな「教えと学び」の仕組みづくり~』 . <https://www.doshisha.ac.jp/news/detail/001-Jhli1.9.html>.
- (12) 同志社大学 (2024). 『学習を支援する AI アシスタント「DAIB」の活用開始について』 . <https://cgle.doshisha.ac.jp/cgle/news/detail/036-53i3SP.html>.
- (13) 文部科学省 (2023). 『大学・高専における生成 AI の教学面の取扱いについて』 . [https://www.mext.go.jp/b\\_menu/houdou/2023/mext\\_01260.html](https://www.mext.go.jp/b_menu/houdou/2023/mext_01260.html).