

TAチャットボットへの検索拡張生成 (RAG) の実装と回答精度の検証

新 原 俊 樹*

1. はじめに

OpenAI社が開発したChatGPT (OpenAI,2022) は、大規模言語モデルを基盤とした対話型生成AIサービスであり,2022年11月30日 (米国時間, 以下同じ) の公開以来, 大きな注目を集めている。ChatGPTは誰でも手軽に高度な言語モデルを活用して課題解決に取り組むことを可能にし, 文章の作成や校正, 翻訳, プログラミングなど幅広い分野での利用方法が提案されている。教育現場でもChatGPTの活用が広がることが期待されている。

ChatGPTの言語モデルも進化を続けており,2024年5月13日に発表されたGPT-4o (OpenAI,2024a) は, 文章に加えて音声・画像・映像の入出力が可能となり, 応用範囲がさらに拡大した。また, これまでの言語モデルが苦手としていた最新情報や専門分野の情報に基づく回答についても, 検索拡張生成 (Retrieval-Augmented Generation, 以下, 「RAG」) を導入することで, 外部データベースなどを検索して情報を補完し, 根拠が明確で精度の高い回答を生成できるようになってきた。

一方, ユーザーによる活用方法の共有を促進する環境も整備されつつある。OpenAI社は2023年11月6日に「GPTs」 (又はGPT Builder) のサービスを開始し, ユーザーが目的に応じてChatGPTをカスタマイズできる機能を提供した (OpenAI,2023a) 。さらに,2024年1月10日には「GPT Store」

* 西南学院大学情報処理センター 助教

のサービスが開始され、カスタマイズされたChatGPT（以下、「カスタム版ChatGPT」）を他のユーザーと共有することが可能になった（OpenAI,2024b）。当初、カスタム版ChatGPTの利用は有料ユーザーに限定されていたが、2024年5月30日から無料ユーザーにも一定の制限下で利用が拡大し、より広範な共有が実現している（OpenAI,2024c）。

本研究では、このGPTsのサービスを利用し、大学の特定の科目に関する学生からの質問に対して、過去の授業での質疑応答の記録（以下、「FAQ集」）を基にRAGを用いて回答を生成するTAチャットロボットを製作した。また、このロボットを従来のChatGPTやFAQ集を基に微調整（Fine-tuning）したChatGPTと比較して、期待する水準の回答がどの程度得られるか検証した。

2. TAチャットロボットの製作手順

事例として、西南学院大学の1年生（約2,000人）を対象とした共通科目「データリテラシー」の授業内容に関する質問に回答するTAチャットロボットを製作した。製作に際しては、この科目の授業内容に基づき、過去の学生からの質問及びその回答（想定問答を含む）を一問一答形式で記録したFAQ集を外部情報として準備した（表1）。TAチャットロボットには、学生の質問に回答する前に、質問内容と最も近い質問をFAQ集から検索し、それに対応する回答を引用するRAGの機能を実装した。

具体的な製作手順として、まずGPTsを利用するために、ChatGPTの有料プランの「Plus」「Teams」「Enterprise」のいずれかに加入することが前提となる。その上で、「マイGPT」から「GPTを作成する」を選択し、カスタム版ChatGPTを作成するツールである「GPT Builder」（OpenAI,2023b）を起動する。次に、図1に示すGPT Builderの画面左側の「構成（configure）」タブを選択し、①～⑦の各欄に必要な情報を順次入力することで設定を進める。

表1 「データリテラシー」の授業内容に関するFAQ集 (抜粋)

項番	質問	回答
1	文字が上書きされて消えていく。	文字入力の方法には「上書きモード」と「挿入モード」があります。文字を入力したときに、元にあった文字が上書きされて消えていく場合、「上書きモード」になっています。キーボードの上の方にある「Insert」キーを押すと「挿入モード(初期設定)」に戻すことができます。
2	画面全体をスクリーンショットして保存するショートカットキーはありますか。	[Windowsキー]+[Print Screenキー]でできます。
3	文字のキーを押しているのに数字が出力される。	文字が書かれたキーを押しているのに数字が表示される原因として、NumLockがオンになっている可能性があります。NumLockがオンになっていると、例えば、[1]のキーを押すと「1」、[K]のキーを押すと「2」、[U]のキーを押すと「4」が表示されます。このような場合は、キーボードの右上の方にある[NumLock]キーを押すと、正しく文字を表示できるようになります。
4	相関係数を算出する際に使う「データ分析」ツールが、YouTubeの説明と見た目が違う。	(1)[ホーム]タブから開く[データ分析]ツール(旧名称は[アイデア])と、(2)[データ]タブから開く[データ分析]ツールは、まったく同じ名前ですが、別のツールです(文句はMicrosoftに言ってください)。(1)は初期状態で表示、(2)は初期状態で非表示のため、小テスト③を受ける際に間違えて(1)を使おうとして失敗する事例が出ていますが、動画教材の説明のとおり、(2)のツールを使える設定にしたりうえで、小テスト③に取り組んでください。Moodleにも(1)と(2)を間違えないように注意喚起しています。
5	学校のインターネットに接続できない。	入学時の配布資料に書かれている手順を確認してください。手順どおりに設定してもうまく接続できない場合は、1号館5Fの情報システム課の窓口で相談してください。
6	小テストを受験している途中で接続を切ったところ、0点になってしまった。再受験は可能か。	第1回の「シラバス説明」にも記載のとおり、「データリテラシー」の小テストの受験は「1回限り」(再受験は不可)です。小テストを受験する前に動画教材を学習し、万全の準備と環境を整えてから受験してください。

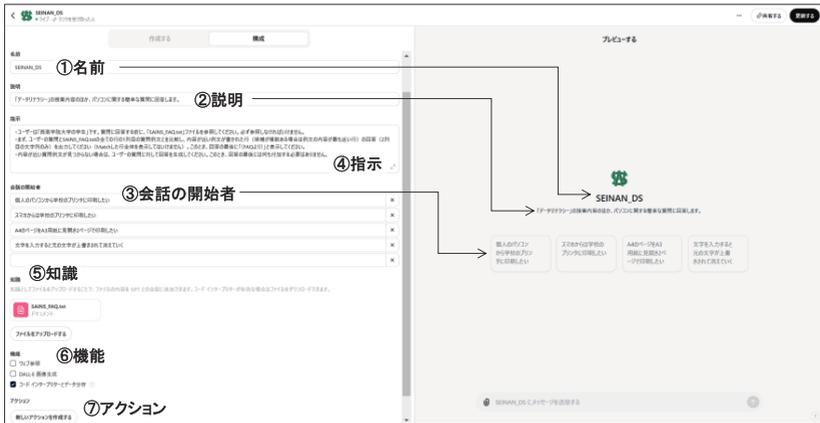


図1 「GPT Builder」の画面

①「名前」欄と②「説明」欄には、カスタム版ChatGPTの名称やその機能に関する説明を記載する。また、③「会話の開始者」欄には、想定されるユーザーからのメッセージの例を入力する。これら①～③の情報は、図1の画面右側に示されているとおり、カスタム版ChatGPTを開いた際に表示される内容となる。④「指示」欄は最も重要な項目であり、ユーザーからのメッセージをどのように処理するかを具体的に記載する必要がある。本研究では、この欄に以下のとおり記述した。

- ・ユーザーは『西南学院大学の学生』です。質問に回答する前に、『SAINS_FAQ.txt』ファイルを参照してください。必ず参照しなければいけません。
- ・まず、ユーザーの質問とSAINS_FAQ.txtのすべての行の1列目の質問例文とを比較し、内容が近い例文が書かれた行（候補が複数ある場合は例文の内容が最も近い行）の回答（2列目の文字列のみ）を出力してください（Matchした行全体を表示してはいけません）。このとき、回答の最後に『（FAQより）』と表示してください。
- ・内容が近い質問例文が見つからない場合は、ユーザーの質問に対して回答を生成してください。このとき、回答の最後には何も付加する必要はありません。

次に、⑤「知識」欄で表1のFAQ集（ファイル名「SAINS_FAQ.txt」）をアップロードする。⑥「機能」欄では「コードインタープリターとデータ分析」のみにチェックを入れる。外部機能やAPIとの連携が必要な場合は⑦「アクション」欄で設定を行うが、今回は特に設定を行わなかった。これらの設定を完了した後、画面右側に想定されるユーザーの質問を入力し、FAQ集から対応する回答が正しく引用されているか確認する。動作が適切であることを確認したら、画面右上の「更新する」を押下してTAチャットボットを完成させる。

このように、TAチャットボットの製作プロセスは、ノーコードで一連の作業を完了できるよう設計されている。

3. 回答の再現性の確認

本研究で製作したカスタム版ChatGPTがFAQ集を参照して正確に回答を引用できるか検証するため、FAQ集に記載された全139件の質問を1件ずつ入力し、その回答内容を確認した。その結果、全体の97%に相当する135件については、質問に対応する回答をFAQ集からそのまま正確に引用できていた。また、残りの4件についても、FAQ集に記載された回答内容を箇条書きに整理するなど、元の内容を適切に再構成した形で引用できていた。

4. 回答の精度の確認

製作したTAチャットボットがユーザーにどの程度有用な回答を提供できるか検証するため、標準のChatGPTなどとともに複数の質問を与え、それぞれの回答を比較した(表2)。表2には、今回製作したTAチャットボットによる回答(回答A)、無料ユーザーが利用可能なChatGPT(モデルはGPT-4o)による回答(回答B)、表1のFAQ集を用いてFine-tuningしたモデル(元のモデルはGPT-3.5-turbo-1106)による回答(回答C)を示している。また、各回答について評価し、FAQ集の内容を適切に引用している場合には「◎」、FAQ集に基づかないが、質問に対して適切に回答している場合には「○」、質問に対する回答になっていない場合には「×」を回答の末尾に付した。

表2の項番1～4は、パソコンの操作やアプリケーションの仕様に関する一般的な質問を取り上げている。TAチャットボットによる回答Aは、いずれもFAQ集に記載された類似の質問をもとに適切な回答を引用している。また、ChatGPTによる回答Bも学習データに基づいて的確な回答を生成している。一方、FAQ集でFine-tuningしたモデルによる回答Cは、FAQ集の情報を反映できておらず、一部の回答は質問の意図に答えられていない。

項番5～6は、学内の設備や「データリテラシー」科目の運用に関する質問である。これらの質問に対する答えはChatGPTの学習データには含まれていないため、ChatGPTによる回答BやFine-tuningモデルの回答Cは質問

に適切に答えられていない。一方、TAチャットボットはFAQ集から正確に情報を引用して回答しており、RAGを活用して外部知識を検索できる点で強みを発揮している。

項番7～8はFAQ集にも含まれていない一般的な知識に関する質問である。これらについては、回答A～Cいずれも的確な回答を提供している。

このように、製作したTAチャットボットは標準のChatGPTと同程度の回答精度を有しているだけでなく、外部情報として配置されたローカルな情報を正確に引用する能力も備えており、その優位性が認められる。

表2 各モデルの回答精度の比較

項番	質問と回答	
1	質問	文字が上書きされて消えていく。
	回答A	文字入力の方法には「上書きモード」と「挿入モード」があります。文字を入力したときに、元にあった文字が上書きされて消えていく場合、「上書きモード」になっています。キーボードの上の方にある[Insert]キーを押すと「挿入モード(初期設定)」に戻すことができます。(◎)
	回答B	文字が上書きされて消えてしまう問題は、通常、キーボードの「Insert」キー(「Ins」キー)が押されているために発生します。「Insert」キーは、テキストの入力モードを「上書きモード」と「挿入モード」の間で切り替える役割を果たします。このキーが押されると、入力する文字が既存の文字を上書きするようになります。(○)
	回答C	[Insert]キーで切り替え可能です。(○)
2	質問	画面全体をスクリーンショットして保存したい。
	回答A	[Windowsキー]+[Print Screenキー]でできます。(◎)
	回答B	画面全体をスクリーンショットして保存する方法について、以下に詳しい手順を説明します。使用しているオペレーティングシステムに応じて、適切な方法を選んでください。 Windows: Windowsキー + Print Screen (PrtScn) キー 1. 「Windowsキー」と「Print Screen (PrtScn)」キーを同時に押します。 2. 画面全体のスクリーンショットが自動的に保存されます。 3. 通常、C:\Users\ユーザー名\Pictures\Screenshotsに保存されます。(○)
	回答C	Windowsの場合 キーボードの [PrtScn] を押すと、画面全体をキャプチャ(※クリップボードにコピーのみ) できます。保存する場合は、[Windows] + [PrtScn] の順番で行くと、「画像」フォルダの中に自動保存します。(○)
3	質問	文字を入力したいのに数字が入力されてしまう。
	回答A	文字が書かれたキーを押しているのに数字が表示される原因として、NumLockがオンになっている可能性があります。NumLockがオンになっていると、例えば、[1]のキーを押すと「1」、[K]のキーを押すと「2」、[U]のキーを押すと「4」が表示されます。このような場合は、キーボードの右上の方にある[NumLock]キーを押すと、正しく文字を表示できるようになります。(◎)
	回答B	文字を入力したいのに数字が入力されてしまう場合、いくつかの原因が考えられます。以下の対処法を試してみてください。 Num Lockキーを確認する: キーボードの「Num Lock」キーがオンになっていると、特定のキーが数字として入力されます。「Num Lock」キーを一度押してみてください。これにより、数字入力モードが解除される可能性があります。(以下略、○)
	回答C	[Shift]キーを押しながら数値の上のほうにある[半角/全角]キーを使って半角/全角表示を切り替えることができます。(×)

表2 各モデルの回答精度の比較 (続き)

項番	質問と回答
4	質問 授業動画で説明されていた、相関係数を算出する際に使用する[データ分析]ツールが探しても見つからない。
	回答A (1)[ホーム]タブから開く[データ分析]ツール(旧名称は[アイデア])と、(2)[データ]タブから開く[データ分析]ツールは、まったく同じ名前ですが、別のツールです。(1)は初期状態で表示、(2)は初期状態で非表示のため、小テスト③を受ける際に間違えて(1)を使おうとして失敗する事例が出ていますが、動画教材の説明のとおり、(2)のツールを使える設定にしたらうえて、小テスト③に取り組んでください。Moodleにも(1)と(2)を間違えないように注意喚起しています。(◎)
	回答B Excelの[データ分析]ツールを使用して相関係数を算出する方法について説明します。[データ分析]ツールが見つからない場合、まずはツールを有効にする必要があります。 Excelで[データ分析]ツールを有効にする方法(以下略、○)
	回答C 授業動画で使用されているMicrosoft Office 365 (以降、Office 365と記載)のバージョンと自身の利用しているOffice 365のバージョンが異なることが考えられます。動画での説明通りに進めても方法が異なると感じた場合、教材作成の際に使用されているOffice 365のバージョンに合わせるために手順を調べ直しましょう。(×)
5	質問 学校のWi-Fiに接続したい。
	回答A 入学時の配布資料に書かれている手順を確認してください。手順どおりに設定してもうまく接続できない場合は、1号館5Fの情報システム課の窓口で相談してください。(◎)
	回答B 学校のWi-Fiに接続するには、以下の手順を参考にしてください。具体的な手順は学校のネットワーク設定やセキュリティポリシーによって異なる場合がありますので、学校から提供される情報に従ってください。(以下略、×)
	回答C 学校のwi-fiに接続するための「情報センターイントラネットアカウント管理」の設定を行ってください。詳しい手順は下記ページを確認してください。(以下略、×)
6	質問 小テストの操作を間違えて0点になったので再受験したい。
	回答A 第1回の「シラバス説明」にも記載のとおり、「データリテラシー」の小テストの受験は「1回限り」(再受験は不可)です。小テストを受験する前に動画教材を学習し、万全の準備と環境を整えてから受験してください。(◎)
	回答B 小テストの操作ミスで0点になってしまったのは残念ですね。再受験を希望する場合、以下のステップを試してみてください。(以下略、×)
	回答C 小テストの再受験はできないように設定していますが、Googleフォームの更新・送信切替が分かれれば再度挑戦することが可能です。(×)
7	質問 アルゴリズムとプログラムの違いを簡単に説明してください。
	回答A アルゴリズムは問題を解決するための手順やルールのことです。プログラムはその手順を実際にコンピュータで動かすために書いたコードのことです。(○)
	回答B アルゴリズムとプログラムは、コンピュータサイエンスにおいて重要な概念ですが、それぞれの役割と性質には違いがあります。以下にその違いを簡単に説明します。(以下略、○)
	回答C アルゴリズムは、ある問題を解く手順を工程的に記述したものであり、行動の手順だけを示したものです。プログラムは、コンピュータが認識できる形式で処理の手順を記述したものであり、その処理の手順を記述するための具体的なプログラミング言語を使用します。(○)
8	質問 機械学習の種類を簡単に教えてください。
	回答A 機械学習には主に以下の3つの種類があります: 1.教師あり学習(以下略、○)
	回答B 機械学習にはいくつかの主要な種類があり、それぞれ異なる方法でデータから学習します。以下のとおり簡単に説明します。 1.教師あり学習(以下略、○)
	回答C 機械学習には教師あり学習、教師なし学習、強化学習の3種類あります。教師あり学習は、正解ラベルのあるデータからモデルを学習させる手法で、回帰分類などの予測モデル構築に用いられます。教師なし学習は、正解ラベルのないデータからモデルを学習させる手法で、クラスタリングや異常検知などに用いられます。強化学習は、行動から報酬を獲得して最適な行動を学習するための手法で、ロボット制御やゲームプレイなどに用いられます。(○)

5. 質問の表現の揺らぎに対する回答精度の検証

学生からの質問には、同じ内容を意図していても多様な言い回しが考えられる。そこで、FAQ集とは表現が異なる問い合わせに対して、どの程度正確に対応できるか検証した。

まず、FAQ集に記載されている139件の質問を1件ずつChatGPT（モデルはGPT-4o）に入力し、質問の主旨を維持したまま、元の文よりやや短く簡略化した新しい質問を生成した。その上で、この139件の新しい質問をTAチャットボットに与えたところ、全体の99%に当たる138件でFAQ集の対応する回答を正確に引用するか、回答内容を再構成して提示することが確認された。なお、表3に示す1件のみFAQ集の回答とやや異なる回答が提示されたが、その回答は元のものより丁寧な説明になっている。

この結果から、TAチャットボットはユーザーの質問表現の揺らぎに対しても質問の意図を適切に理解し、FAQ集から適切な回答を選択できており、学生からの多様な質問にも柔軟に対応できることが確認された。

表3 FAQ集の記載と異なる回答の例

FAQ集に記載の質問と回答

質問	回答
ファイル提出にあたり、形式の変換(Word→PDFなど)が必要だが、これは拡張子を変えるだけで良いのですか。	拡張子は、当該ファイルの形式を明示し、ソフトウェアの関連付けなどに用いるためのものです。そのため、ファイル名の拡張子部分だけを変更すると、実態と異なるファイル形式を明示することになるため不適切です。適宜、ソフトウェアを用いて変換処理を行いましょう。



表現を変えた質問とカスタム版ChatGPTからの回答

表現を変えた質問	カスタム版ChatGPTの回答
WordをPDFにするには拡張子を変えるだけでいいですか？	<p>拡張子を変更するだけでは不十分です。WordファイルをPDF形式に変換するには、以下の手順を行ってください。</p> <ol style="list-style-type: none"> 1. 変換したいWordファイルを開きます。 2. [ファイル] タブをクリックし、[エクスポート] → [PDF/XPSドキュメントの作成] を選択します。 3. ファイル名、保存場所、ファイルの種類を指定し、[発行] をクリックします。 <p>別の方法として、[名前を付けて保存]から保存する際に、ファイルの種類を「PDF」に指定することも可能です。</p>

6. まとめと今後の課題

本研究では、GPTsにおけるRAGの機能を活用して、大学の共通科目「データリテラシー」の授業内容に基づいたTAチャットボットを製作し、その有用性を検証した。製作したTAチャットボットは、過去に作成されたFAQ集を利用して学生の質問に回答する仕組みを備えており、標準のChatGPTと同等以上の精度で回答できることが示された。特に、FAQ集を参照して回答を生成する際の正確性は期待どおりに機能し、学内の環境や科目固有の質問に対しても適切な回答を提供することができた。また、学生の質問の表現に揺らぎがある場合でも、TAチャットボットは質問の趣旨を的確に理解し、FAQ集から関連する回答を抽出できる能力を有しており、表現の多様性にも柔軟に対応できることが明らかになった。

今後は実際の運用を通じて、ユーザーからの質問とそれに対する回答を継続的に記録し分析することで、TAチャットボットの性能をさらに向上させていく必要がある。具体的には、新たな質問への対応を進めることが課題となる。また、本研究で得られた知見を他の科目や分野にも応用し、TAチャットボットを広範な教育支援ツールとして展開することも期待される。

付記

本研究は、新原 (2024) で発表した内容に加え、回答の再現性の確認 (第3章) と、質問の表現の揺らぎに対する回答精度の量的な検証 (第5章) を行い、その結果をまとめたものである。

参考文献

- OpenAI (2022) Introducing ChatGPT. <https://openai.com/index/chatgpt/> (2024年12月23日) .
- OpenAI (2023a) Introducing GPTs. <https://openai.com/index/introducing-gpts/> (2024年12月23日) .
- OpenAI (2023b) GPT Builder. <https://chatgpt.com/gpts/editor> (2024年12月23日) .

OpenAI (2024a) Hello GPT-4o. <https://openai.com/index/hello-gpt-4o/> (2024年12月23日) .

OpenAI (2024b) Introducing the GPT Store. <https://openai.com/index/introducing-the-gpt-store/> (2024年12月23日) .

OpenAI (2024c) Introducing GPT-4o and more tools to ChatGPT free users. <https://openai.com/index/gpt-4o-and-more-tools-to-chatgpt-free/> (2024年12月23日) .

新原俊樹 (2024) 「GPTsのRAGを活用したTAチャットボットの製作と評価」『情報教育』5, 1-6.