

# 生成系 AI を利用した小学生向けの個別学習支援システムの検討と実装

榎田透馬<sup>1</sup> 寺澤卓也<sup>1</sup>

**概要：** ChatGPT に代表される生成系 AI の登場以降、様々な場面でそれらが活用されるようになった。生成系 AI はプロンプトと呼ばれる指示がどのくらい詳細で適切かによって生成結果の質が変化する。教育の現場において ChatGPT を活用する場合、特に小学生にとっては適切なプロンプトを考え ChatGPT に与えることが難しい。そこで本研究では、児童生徒一人一人のアシスタントとして ChatGPT を活用するためにプロンプト作成の補助を自動的に行って回答を引き出すシステムを提案する。第一段階の実装として Web アプリケーション化を行い、評価実験を実施した結果、本手法の有効性が示された。また、考察とともに今後のシステムの拡張計画と展望を示す。

**キーワード：** 生成 AI プロンプト作成 個別学習支援

## An individual learning support system for primary school students using generative AI

TOMA KUSHIDA<sup>1</sup> TAKUYA TERASAWA<sup>1</sup>

### 1. はじめに

2019 年の学習指導要領改訂[1]により、ICT 機器を活用した個別最適な学びが推進され始めた。個別最適な学びは、指導の個別化と学習の個別化の二つで構成されている。指導の個別化は、生徒一人一人の特性や学習速度に応じて指導方法や学習環境を調整することである。一方、学習の個別化は、生徒の興味関心がある事柄に応じた学習活動や課題を提供し、学習内容が最適となるよう生徒自ら調整することである。個別最適な学びを推進するにあたって機器の普及が不可欠であることから、GIGA スクール構想[2]が打ち出された。GIGA スクール構想とは、全ての生徒に一人一台のデジタル端末を配備し、高速大容量のインターネット環境を学校に整備するという政策である。2022年3月の時点でコンピュータ配備率は平均 109.2%になり、この構想の目的は達成されている。しかし、ICT 機器の活用に関する課題が依然として残っている。2021 年文部科学省の GIGA スクール構想に関する課題の調査[3]によると、学校の学習指導での活用と教員の ICT 活用指導力に課題があるとする回答が多くあった。構想による端末の配備だけでは、教員が ICT 機器を活用した指導の知識やノウハウを持っていないため不十分である。一方、生成系 AI が急速に発達しており、各方面でその活用法が探られている。生成系 AI には注意を払うべき点もあるが、上手に活用すれば個別最適な学びを実現できる可能性がある。

近年開発されている AI の中で、ChatGPT[4]や Gemini[5] に代表される生成系 AI の発展が目覚ましい。これらのサービスはチャット形式で行われ、AI に対してプロンプト

として質問を送ることで回答を得る仕組みになっている。非常に便利なサービスなため、簡単に扱うことが出来るが、意図した回答を得るためには適切なプロンプトを作成するスキルが必要である。

本研究では、個別最適な学びに対応するため、生徒一人一人が持っている能力にかかわらず、生成系 AI をアシスタントとして活用できるシステムの検討と実装を行う。システムの利用は授業や家庭学習時を想定し、プロンプトを作成するスキルを持っていない小学生でも、意図した回答を得られるような仕組みを構築する。

生成系 AI は同じ質問に対して異なる回答を返すことがある。しかし、初等教育における学習という観点からは、同じ内容の質問に対して質問者や質問タイミングによって異なった回答が生成されることは望ましくない。また、生徒がよく質問を行う問題やその傾向を教員側が容易に把握できる機構も現場の観点から必要である。これらを解決する手段として、システムを Web アプリケーションとして構築する。こうすることで Web アプリケーションサーバーが API 経由で生成系 AI を利用し、質問や回答をデータベース化することもできる。同様の質問の場合データベースから回答を参照することで、回答の同一性を保ち、質問傾向の分析を行うと共に質問の回答生成にかかる時間の軽減を図る。

### 2. 関連研究

生成系 AI に送るプロンプトの作成補助研究はすでに行われている。Li らの研究[6]では、強化学習や教師あり学

<sup>1</sup> 東京工科大学メディア学部  
School of Media Science Tokyo University of Technology

習を用いて独自にトレーニングしたモデルへ、ユーザーが作成したプロンプト文と過去の入力履歴を与えることで、プロンプト文を最適な文へ自動的に修正するシステムを開発した。ユーザーが作成したプロンプトを要約し、質問履歴やワードの学習によって最良であるとしたキーワードを統合することで、元のプロンプトや強化学習等で最適化されたプロンプトよりも優れた結果が得られている。この研究では事前に作成されたプロンプトに対し最適となるように補正を行っているが、はじめのプロンプトを作成するための補助は行っていない。

Dong の研究[7]では、GPT-4 モデルと RAG[8]技術を用いて質問に対してよりの確に回答する手法を提案している。RAG とは、検索拡張生成 (RAG: Retrieval Augmented Generation) のことで、生成系 AI とは別に、外部にデータベースを作成し、送られてきた質問を基にデータベースで検索を行い、関連する項目を情報として与えることで、質問に対して関連度が高く正確な回答の生成を行えるようにする技術である。この研究では、利用者が事前に資料をアップロードすることでデータベースを作成し、そのデータに基づいた回答を GPT-4 モデルによって生成している。RAG を用いることで、正確性や関連性の向上や資料の正確な引用が見られたとしているが、この手法の問題点として、質問が曖昧または広すぎる場合に適切な回答を生成できない可能性がある指摘している。

### 3. 提案手法

本章では、提案手法の概要と使用する生成系 AI について述べる。

#### 3.1 提案手法の概要

図 1 に提案手法の全体図を示す。本手法では、質問者が意図した回答を生成系 AI から得られるように、曖昧な質問文を生成系 AI によってプロンプトに変換し、そのプロンプトを用いて生成系 AI に回答させる。このプロンプト作成の補助を行うことにより、生成系 AI の利用に関してのハードルを下げ、質問者が意図した回答を得られるようにすることを目的としている。

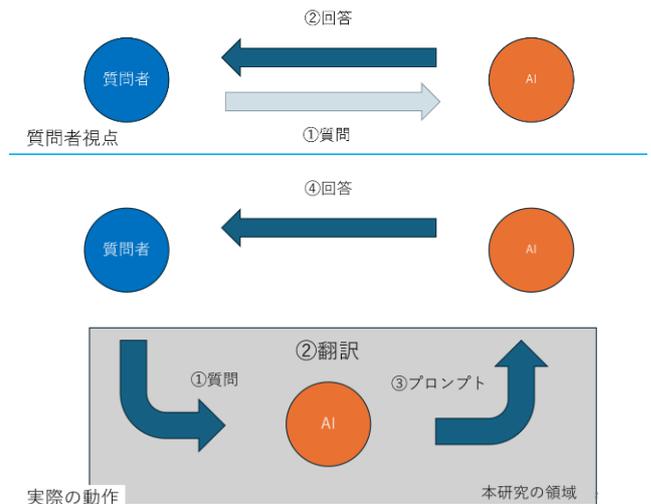


図 1 提案手法の全体図

#### 3.2 システムの想定利用者と科目

第 1 章で述べた通り、本研究は小学生を対象としている。小学生は大人よりも言語能力が低く、知識量も少ないため、質問内容の明確化が難しい。そのため、生成系 AI に誤解の無いプロンプトを与えることが難しいことから、本手法での補助が最も有効に機能すると考えた。また、科目を算数に限定して評価を行う。算数は言語や文化による差が少ないため、生成系 AI の学習データ量が多く、評価を定量的に行いやすいことから、算数に限定し実験を行った。

#### 3.3 各生成系 AI の概要

使用する生成系 AI を選定するため比較調査を行った。性能の良い GPT-4[9]モデルの使用や学習データ量を考慮し、Copilot[10]、ChatGPT(GPT-4)、Gemini、Perplexity[11]の 4 つを調査対象とした。また、調査は 2023 年 12 月 15 日に行った。3.3.1 項に各生成系 AI についての概要を、3.3.2 項に比較調査結果を示す。

##### 3.3.1 生成系 AI の比較調査

Copilot は Microsoft が開発した、GPT-4 を大規模言語モデルとした AI サービスである。Microsoft が提供しているサービスとの連携が可能で、検索エンジンの Bing と連携し、Web 検索を用いて最新の情報から回答を生成することができる。

ChatGPT は OpenAI が開発した、GPT-4 を大規模言語モデルとした AI チャットボットである。AI チャットボットサービスの先駆けであり、インターネット上の大量のデータを学習することで、高い回答精度を実現している。中でも GPT-4 モデルは、現在多くの AI サービスの大規模言語モデルとして用いられている。

Gemini は Google が開発した、Gemini を大規模言語モデルとした AI チャットボットである。2024 年 2 月 8 日に Google Bard から Gemini に名称が変更された。大量のデータを用いて学習され、性能テストの結果[12]、32 項目中 30

項目でGPT-4モデルよりも高い性能を示したとされている。

PerplexityはPerplexity AIが開発した、GPT-4を大規模言語モデルとしたAIチャットボットである。Copilotと同様に、質問を基にインターネットから情報を収集し回答を生成する。

### 3.3.2 比較調査結果

図2に、比較調査に用いたプロンプトと小学生からの想定される質問3種を示す。想定される質問は上から4, 5, 6年生の問題[13]に相当し、みかんのような名詞を含むもの、速さのような抽象的な概念を含むもの、水槽問題のような文章題で複雑な計算が必要なものという理由でそれぞれ選定した。まず、名詞を含むことにより、算数の問題以外の事柄についての説明をするかのテストを行った。次に、速さのような抽象的な概念を、小学生が理解できる内容と語彙で説明できるかのテストを行った。最後に、水槽問題のような文章題で複雑な計算が必要なものについて、曖昧な質問から説明文を生成できるかのテストを行った。このプロンプトと質問を生成系AIに送り、以下を調査した。

- ・プロンプトの指示に従った文章生成を行えるか
- ・算数の質問として答えられるか

みかんの質問(1番)に対する回答結果を図3, 図4, 図5図6にそれぞれ示す。また、wordでカウントしたそれぞれの文の文字数を文末に記した。調査の結果、Copilot, Gemini, Perplexityにおいて、プロンプトの指示に従えていないものと質問に答えられていないものがあった。Copilotでは、図3のように算数の質問として答えていないことや2番の質問でスカラーやベクトルなどの小学生が未習である言葉を用いての説明と文字数違反があった。Geminiでは、小学生が理解しやすい語彙での説明だったが、図5のような文字数違反がすべての質問であった。Perplexityでは、3番の質問のみ文字数違反があった。調査をして問題がなかったChatGPTと問題が少なかったPerplexityを比べた時、APIによるカスタマイズ性がChatGPTの方が高く、アプリケーション化を行いやすいことから、今回はChatGPTを使用する。

プロンプト	想定される質問
以下の質問について、小学生が理解できる内容と言葉遣いで、300文字以内で説明してください。 算数の質問：???	1. みかん1個分の値段はどうですか？ 2. 速さって何？ 3. 水槽をいっぱいにするのにかかる時間がよくわからない

図2 比較調査に用いたプロンプトと質問

もちろんです！みかんの値段は、スーパーやお店によって異なりますが、一般的には1個あたり20円から40円くらいです。たとえば、小袋で10個のみかんが200円から400円なら、1個あたりの値段はこの範囲になりますね。みかんは美味しく栄養たっぷりなので、食べるのが楽しみです！(137文字)

図3 Copilot 生成文

質問：みかん1個分の値段はどうですか？

みかんとたくさん買った時の全部の値段と、買ったみかんの個数を知っていれば、1個分の値段がわかるよ。全部で払ったお金をみかんの個数で割るんだ。例えば、みかんと10個で1000円で買ったなら、1000円を10で割って、みかん1個あたりの値段は100円になるんだよ。(129文字)

図4 ChatGPT 生成文

質問：みかん1個分の値段はどうですか？

お店に並んだみかん、どれも美味しそう！でも、1個だけ買いたいときは、いくら払えばいいのかな？  
みかん1個分の値段を出すには、みかんの値段と個数を関係する魔法を使うんだ！  
1.みかんの値段を知る  
まず、お店の人に聞こう。「みかんはいくらですか？」って。すると、例えば「1kgで300円」と教えてくれるね。  
2.みかんの個数を数える  
次に、1kgの中のみかんがいくつ入っているか数えよう。10個、20個…数えてみると、1kgには25個みかんが入っていたね。  
3.魔法を使う！  
そして、魔法の言葉「300円 ÷ 25個」を唱えると…1個分の値段がわかる！  
300円 ÷ 25個 = 12円  
なんと、みかん1個は12円だったんだ！  
このように、みかんの値段と個数を知れば、魔法を使って1個分の値段を簡単に調べることができるんだ。  
お店でみかんを見かけたら、ぜひこの魔法を使って、1個分の値段を計算してみてね！  
(383文字)

図5 Gemini 生成文

質問：みかん1個分の値段はどうですか？

みかん1個の値段を求めるには、みかんの総額をみかんの個数で割ればよいです。例えば、みかんが100円で10個入っている袋がある場合、1個あたりの値段は100円÷10個で計算できます。したがって、みかん1個の値段は10円となります。  
(114文字)

図6 Perplexity 生成文

質問：みかん1個分の値段はどうですか？

### 3.4 プロンプト作成の流れと例

3.3.2の調査結果に基づき、ChatGPTを使用して質問文から適切なプロンプトを作成する実験を2024年1月19日に行った。与えられた質問から回答文を生成する流れを図7に示す。また、「みかん1個分の値段はどうですか？」という質問が送られてきたと想定し、図7の流れに沿って実際に作成したプロンプトと回答文をそれぞれ図8, 図9, 図10に示す。筆者が作成したプロンプトを黒字、ChatGPTが生成したものを赤字、生成されたプロンプトを組み込んだ部分を緑字とする。

まず、送られてきた質問からその内容の抽出を行う(図8)。ここで送られてくる質問は抽象的な短い文と想定し、少ない情報から説明文を生成するためのプロンプトが必要となる。Liら[14]の、文章からキーワードを抜き出し、ヒントとして与えることで回答の方向性を定める **Directional Stimulus Prompting** 技法を参考に、トピックと疑問点を分類することで、生成文の内容を決定する。Liらの研究では、長文テキストの読解のためのヒントとして、文章に含まれる要素をヒントとして渡すことで、回答の精度を上げている。これを応用し、要素となるヒントをユーザーが入力する質問から抽出することで、回答の方向性を決定させる。次に、分類を行ったトピックと疑問点をプロンプトに組み込み(図9)、Kojimaら[15]の **Step by Step** 法を参考にステップ毎に文章を生成する(図10)。ステップ毎に回答を生成させることで、全体を一度に生成する際よりも回答の精度が高くなることが Kojima らの研究で分かっている。これを用いることで、人間が説明する時と同じ順序になると共に、段階的に文章を生成することによる回答精度の向上を図る。



図7本システムにおける回答文生成のための流れ

これから小学生からの算数の質問が送られます。その内容を以下の条件に沿って特徴を抽出し分類してください。質問は、小学生「円の面積の求め方が分かりません」の形で送られます。

#トピックは何 例：分数の問題、面積の問題  
#疑問点は何 例：時速の求め方が分からない、計算方法が分からない

小学生「みかん1個分の値段はどうですか?」  
↓  
トピックは何 トピック：「個別の価格計算問題」  
疑問点は何 疑問点：「みかん1個の値段の計算方法」

図8 提案手法を用いた質問内容の抽出  
質問：みかん1個分の値段はどうですか?

以下のプロンプトは、小学生からの算数の質問と、トピックと疑問点を抽出したものです。これを使用して、小学生の疑問を解決する文章を生成してください。ただし、学習の道筋を教えるための文章である必要があります。ステップ毎に解き方を分割してください。また、「次をお願いします。」とチャットを返すごとに、ステップを送ってください。

小学生「みかん1個分の値段はどうですか?」  
トピックは何 トピック：「個別の価格計算問題」  
疑問点は何 疑問点：「みかん1個の値段の計算方法」  
↓  
**ステップ1: 問題を理解する**  
まずは、問題をしっかりと理解しましょう。この問題では、みかんのまとめ買いの値段が与えられていると思います。例えば、「みかんが10個で600円です。1個あたりの値段はいくらですか?」のような問題です。

図9 提案手法を用いた説明文の生成(ステップ1まで)  
質問：みかん1個分の値段はどうですか?

**ステップ2: 必要な情報を書き出す**  
次に、問題から必要な情報を書き出します。この例では、  
・みかんの個数(10個)  
・みかんの合計価格(600円)  
が必要な情報です。

**ステップ3: 1個あたりの値段を計算する**  
問題の情報をもとに、1個あたりの値段を計算しましょう。これは、「みかんの合計価格」を「みかんの個数」で割ることで求められます。計算の式は次のようになります：  
$$1 \text{ 個あたりの値段} = \frac{\text{みかんの合計価格}}{\text{みかんの個数}}$$
  
例題では、600円を10個で割ります。したがって、  
$$1 \text{ 個あたりの値段} = \frac{600 \text{ 円}}{10 \text{ 個}} = 60 \text{ 円}$$
  
となります。

**ステップ4: 答えを確認する**  
最後に、計算した答えが妥当かどうかを確認しましょう。60円が1個あたりの価格なので、10個で600円になることを再度確かめます。60円 × 10個 = 600円であるため、答えが正しいことがわかります。

図10 提案手法を用いた説明文の生成(ステップ2~4)  
質問：みかん1個分の値段はどうですか?

#### 4. 評価

3章で示した本システムについて、曖昧な質問から質問者が意図した回答を得られるか、調整を加えていない既存モデルと比較してより良い回答を得られるかを検証するために Web アプリケーションを作成し、2024年6月1日~12日に評価実験を行った。

##### 4.1 アプリケーション概要

アプリケーションは Python の Web アプリケーションフレームワークである Django[16]を用いて作成し、AWS[17]の仮想サーバーを使用して Web 上に公開した。アプリ利用の様子を図11に示す。実際には A, B のサーバーは同一である。応答モデルには GPT4-turbo の API を使用し、比較のために2種類の Chat Bot を作成した。ユーザーの質問をそのまま ChatGPT に流す Chat Bot を A、本システムを組み込んだ Chat Bot を B とし、質問を基にそれぞれの Chat Bot が同時に回答を生成するように構成した。

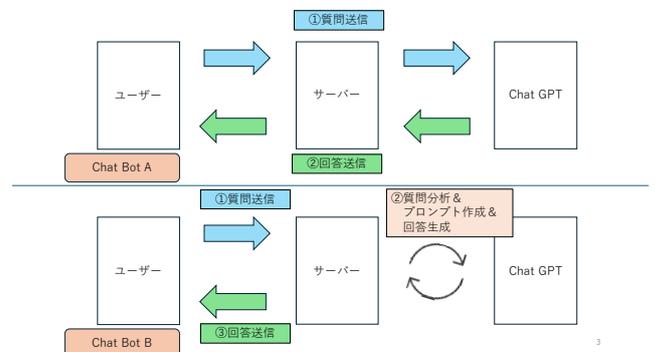


図11 アプリ利用の全体図

## 4.2 実験概要

実験には9名（大学生：8，社会人：1）が参加した。本システムの有効性を測るため、被験者にはわざと主語や術後を抜いた質問や、小学生が行うような語彙での質問を作成してもらった。その質問を基に2種類のChat Botが生成した回答が、それぞれ意図した問題についての説明であったか、またそれぞれの使用感を5件法を用いてアンケートを行った。質問は複数作成可とし、一つの質問ごとにアンケートを取った。実験には、実際の小学生に参加してもらうことが望ましかったが日程として難しかったため、まず大学生・社会人に対して実験を行った。

## 4.3 結果

実験に投入された質問の一覧を表1に示す。質問は計16件集まった。また、質問を基に2種類のChat Botが生成した回答についてのアンケートの平均得点を表2に示す。アンケートは5件法を用いて行い、全く当てはまらないを1点、あまり当てはまらないを2点、どちらともいえないを3点、やや当てはまるを4点、とても当てはまるを5点とした。ユーザーの質問をそのままChatGPTに流すChat Bot Aと本システムを組み込んだChat Bot Bでは、全ての平均においてBの評価が高かった。

Chat Botの使用感についての比較グラフを図12に示す。Bの方が使いやすかったとした回答が14件、どちらとも言えないとした回答が2件、Aの方が使いやすかったとした回答は0件であった。

表1 実験に投入された質問の一覧

No	質問
1	大きい素数はどうやって素数ってわかるの？
2	池のまわりを反対に走る2人が出会うのはいつ？
3	学校から公民館までAくんがBくんを追い越すのはいつ？
4	丸いクリスマスケーキを7つの同じ大きさに分けるには？
5	8を(1/2)で割ると16になるのはなぜ？
6	5角形の一つの角は何度？
7	動く速さってどうやって出すの？
8	かけ算の順番わかんない
9	食塩水の濃さはどう出すの？
10	1から100まで足していったら何になるの？
11	ミカンをみんなに分けたら、ひとり何個分になる？
12	この飛行機は合計でどれくらいの飛んだの？
13	三角形の面積はどう出すの？
14	家から学校まではどのくらいかかる？
15	三角形の大きさはいくつ？
16	池を走ったら出会うのは何分後？

今回、実験に参加したのは大学生以上であったため、小学生らしくない質問も見受けられた。また、文末を曖昧にした質問が多かったが、主語と述語は正しい質問がほと

んどであった。

表2 アンケートの平均得点

質問	Chat Bot A 平均得点	Chat Bot B 平均得点
聞きたかったこと、想定した問題文等に対する確に答えていたか	3.0625	4.6875
生成された文章は分かりやすかったか	3.25	4.625
小学生が理解できる説明であったか	2.5625	4.25

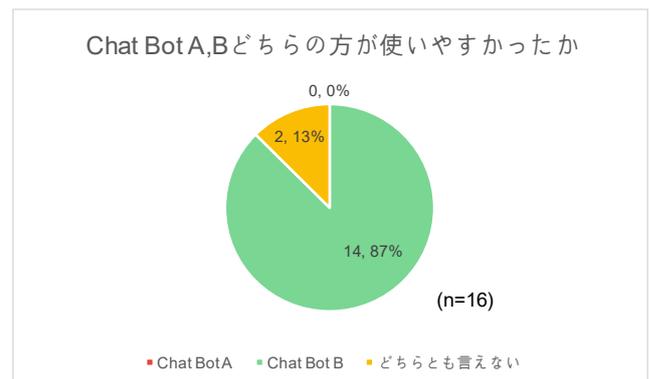


図12 どちらのChat Botが使いやすかったかについてのアンケートグラフ

## 4.4 考察

Chat Bot AよりもBの方が全ての平均得点が高く、個別の質問ごとの得点でも、ほとんどの質問においてBはAと同等かそれ以上の点数であった。この結果から本システムはユーザーの質問をそのままChatGPTに流す場合よりも、曖昧な質問から質問者が意図した回答を生成することができると考えられる。

また、質問1,3,8,14,15においてAとBの得点差が大きかった。その中でも質問3,14,15はAの回答に共通点が見られた。Aは、どの回答も具体的な数値や問題の詳細を質問者に求める内容で、質問についての説明文は生成されなかった。反対にBは、仮の数値を定義して説明文を生成していた。このことから、BはAよりも情報の足りていない質問文から回答を作り出す能力が高く、AとBの得点差が生まれたと考える。

生成された回答に対する評価として、表3にループリッックを示す。一つの質問ごとに着目した場合、小学生が理解できないであろう語彙での説明が多々あった。その例を付録A.1に示す。一般的な歩行速度など小学生が未習の漢字や概念を用いた説明が含まれている。表4にループリッックを基にChat Bot Bが生成した回答を筆者が評価した結果を示す。内容と説明方法に関して、ほとんどの質問で正しく、既習の範囲での説明が行われた。しかし、語彙の難度

に関して、約半数が説明の一部で小学生が理解できないであろう語彙を用いている。実験参加者は大学生と社会人であったため、内容の理解度が高かった可能性が有ることから、語彙の調節を行うと共に、小学生を対象とした評価の実施が必須であると考えられる。

表 3 生成された回答に関するルーブリック

	3	2	1
内容の正確性	正しい内容の説明が行われている	一部正しくない説明が行われている	間違った説明が行われている
語彙の難度	小学生が理解できる語彙で説明が行われている	一部小学生が理解できない語彙で説明が行われている	小学生が理解できない語彙での説明が多々ある
説明方法	小学生が既習の範囲での解き方で説明が行われている	一部小学生が未習の範囲の解き方で説明が行われている	小学生が未習の範囲の解き方で説明が行われている

表 4 ルーブリックに基づく Chat Bot B が生成した回答の評価結果(16 個の質問のうちの該当数)

	3	2	1
内容の正確性	15 個	1 個	0 個
語彙の難度	9 個	7 個	0 個
説明方法	15 個	1 個	0 個

評価にあたり、入力する質問を算数の問題に限定したが、回答文の生成に問題点はほとんど見当たらなかった。問題が起こった質問は、帰納的な考えが必要な問題と「分」のような時間と分数で用いられる多義語が含まれる問題であった。前者は、複数の例を与えて法則性を当てさせる問題で、期待していた法則とは別の法則やこじつけを回答する場合があった。後者は、質問をトピックと疑問点に分解する際に誤った判断が行われたことが原因であった。

評価後、算数以外の質問について、プロンプトの科目名を変更する程度で回答の生成がどのように行われるか確認を行った。回答の精度は質問の内容によって変化があったが、質問内容の分類は算数と同様に行っていた。したがって、他科目についても回答の生成方法を検討することで、このシステムを適用できると考える。また改良点として、再質問が挙げられる。本システムでのユーザーの入力は、最初の質問のみである。質問の内容によって、精度を上げるためには不足した情報を後から追加するなどといった処理が必要になってくる。算数や他科目への適用において回答精度を向上させるためには、生成された回答に対して再度質問する機構を追加する必要があると考える。

## 5. おわりに

本研究では、個別最適な学びに対応するためのシステムとして曖昧な質問から質問者が意図した回答を ChatGPT を用いて生成する手法を提案した。そして Python の Web アプリケーションフレームワークである Django を用いて 2 種類の Chat Bot を作成し評価実験を行った。一方は投入された質問をそのまま ChatGPT に渡し回答を得る Chat Bot であり、もう一方は本研究の手法を適用したものである。評価の結果、本手法を適用した Chat Bot の方が、曖昧な質問文から質問者が意図した回答を生成することができることが示された。

今後の課題として、1 章に記したデータベースの有効活用と質問ごとに適したプロンプトに再調整することが挙げられる。

現状では、質問から生成された回答はデータベースに保存しているが、同じ内容の質問が送られてきた際にデータベースから回答を参照する機能を実装できていない。回答の同一性を担保するためにこの機能の実装は必要であると考えている。しかし、それに伴いデータベース内での検索に時間がかかることが予想される。これを解決する手段としてキャッシュの概念を用いる。通常、生徒の質問は授業の展開によって発生すると考えられる。そのため、複数の生徒が同じ内容の質問を同タイミングで思いつく可能性が高いといえる。キャッシュにおける時間的局所性・空間的局所性の概念から、一定時間に投入された質問と回答の情報をデータベースに記録し、その質問に関連した内容の回答を優先的に検索するようにシステムを構成することで、検索時間の短縮を実現できると考える。

また、質問ごとに適したプロンプトに再調整する必要がある。本システムでは、事前に筆者が作成したプロンプトに基づいて、質問内容の抽出とプロンプトの生成を行っている。そのため、質問内容によって個別に調整することが出来ない状態である。しかし、あらかじめ設定したプロンプトだけでは、どの質問に対しても良い回答を生成することは難しい。よって、質問内容に応じて適切な抽出やプロンプト作成を行うためのプロンプトを作成するシステムを加える必要があると考える。これを解決する手段として、機械学習を用いたプロンプト作成を行う。質問ジャンルによって最適であるプロンプトを学習させることで、それぞれの質問に適したプロンプトで抽出や生成を行うことができ、より精度の高い回答を導き出せる。

これらのシステムの改善を行った後、小学生を対象に評価実験を行うことで本研究の効果を示す。

## 参考文献

- [1] 文部科学省. “平成 29・30・31 年改訂学習指導要領（本文、解説）”. [https://www.mext.go.jp/a\\_menu/shotou/new-cs/1384661.htm](https://www.mext.go.jp/a_menu/shotou/new-cs/1384661.htm), 2019
- [2] 文部科学省. “GIGA スクール構想の実現について”. [https://www.mext.go.jp/a\\_menu/other/index\\_00001.htm](https://www.mext.go.jp/a_menu/other/index_00001.htm), 2023.
- [3] 文部科学省. “GIGA スクール構想に関する各種調査の結果”. [https://www.mext.go.jp/content/20210827-mxt\\_jogai01-000017383\\_10.pdf](https://www.mext.go.jp/content/20210827-mxt_jogai01-000017383_10.pdf), 2021.
- [4] ChatGPT. <https://chat.openai.com/>,
- [5] Gemini. <https://bard.google.com/chat>,
- [6] Cheng Li, et al. “Learning to Rewrite Prompts for Personalized Text Generation” <https://storage.googleapis.com/gweb-research2023-media/pubtools/pdf/f9302b1aacb6b25359be9b1f35e45bd8e12b270.pdf>, 2024.
- [7] Chenxi Dong. “How to Build an AI Tutor that Can Adapt to Any Course and Provide Accurate Answers Using Large Language Model and Retrieval-Augmented Generation”. arXiv preprint arXiv:2311.17696, 2023.
- [8] Patrick Lewis. “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks”. arXiv preprint arXiv:2005.11401, 2020.
- [9] OpenAI. “GPT-4 Technical Report”. arXiv preprint arXiv:2303.08774, 2023.
- [10] Copilot. <https://www.bing.com/search?form=NTPCHB&q=Bing+AI&shoconv=1>,
- [11] Perplexity. <https://www.Perplexity.ai/>,
- [12] Gemini Team, Google. “Gemini: A Family of Highly Capable Multimodal Models”. [https://storage.googleapis.com/deepmind-media/gemini/gemini\\_1\\_report.pdf](https://storage.googleapis.com/deepmind-media/gemini/gemini_1_report.pdf), 2023.
- [13] 文部科学省. “【算数編】小学校学習指導要領（平成 29 年告示）解説”. [https://www.mext.go.jp/component/a\\_menu/education/micro\\_detail/\\_icsFiles/afieldfile/2019/03/18/1387017\\_004.pdf](https://www.mext.go.jp/component/a_menu/education/micro_detail/_icsFiles/afieldfile/2019/03/18/1387017_004.pdf), 2017
- [14] Zekun LI, et al. “Guiding Large Language Models via Directional Stimulus Prompting” arXiv preprint arXiv:2302.11520, 2023.
- [15] Takeshi Kojima, et al. “Large language models are zero-shot reasoners” arXiv preprint arXiv:2205.11916, 2022.
- [16] Django. <https://www.djangoproject.com/>,
- [17] AWS. <https://aws.amazon.com/jp/>,

## 付録

### 付録 A.1 質問 No.14 に対する Chat Bot A と B の回答

[Chat Bot A]  
家から学校までかかる時間は、あなたの住んでいる場所と学校の場所、交通手段によって異なります。歩く場合、自転車を使う場合、公共交通機関を使う場合、または自動車で行く場合など、どの方法を使うかによっても時間は変わります。具体的な距離と移動手段を教えてください。いただければ、もっと詳しく答えることができます。

[Chat Bot B]  
[トピックの分類]  
#トピックは何  
距離と時間の問題  
  
#わからない部分は何  
距離や所要時間の計算方法が分からない  
  
以下のプロンプトは、小学生からの算数の質問と、トピックと疑問点を抽出したものです。これを使用して、小学生の疑問を解決する文章を生成してください。ただし、学習の道筋を教えるための文章である必要があります。ステップ毎に解

きを分割してください。  
小学生家から学校まではどのくらいかかる？  
#トピックは何  
距離と時間の問題  
  
#わからない部分は何  
距離や所要時間の計算方法が分からない[説明文生成時に与えたプロンプト]  
以下のプロンプトは、小学生からの算数の質問と、トピックと疑問点を抽出したものです。これを使用して、小学生の疑問を解決する文章を生成してください。ただし、学習の道筋を教えるための文章である必要があります。ステップ毎に解き方を分割してください。  
小学生家から学校まではどのくらいかかる？  
#トピックは何  
距離と時間の問題  
  
#わからない部分は何  
距離や所要時間の計算方法が分からない  
  
[最終的に生成された文章]  
質問に答えるためには、距離と時間の関係を理解することが大切です。ここでは、家から学校までの距離と、その距離を移動するのにかかる時間を計算する方法を学びましょう。  
  
### ステップ 1: 距離を知る  
まず、家から学校までの距離を知る必要があります。この距離は、地図アプリや学校からの案内、または実際に距離を測ることで知ることができます。例えば、家から学校まで1キロメートルあるとします。  
  
### ステップ 2: 移動速度を考える  
次に、どのくらいの速さで移動するかを考えます。歩く場合、一般的な歩行速度は時速約5キロメートルです。自転車の場合は時速約15キロメートルが一般的です。  
  
### ステップ 3: 時間の計算  
距離と速度が分かれば、時間を計算することができます。時間は「距離 ÷ 速度」の公式で求めることができます。例えば、歩いて学校に行く場合、時間は以下のように計算できます。  
  
- 距離: 1 キロメートル  
- 速度: 5 キロメートル/時  
  
時間 = 距離 ÷ 速度 = 1 キロメートル ÷ 5 キロメートル/時 = 0.2 時間  
  
0.2 時間は分に換算すると、0.2 × 60 分 = 12 分です。つまり、歩いて学校まで 12 分かかるといことになります。  
  
### ステップ 4: 結果の確認  
計算した時間が妥当かどうかを考えてみましょう。もし不安な場合は、実際にタイマーを持って歩いてみると良いでしょう。また、異なる速度で計算してみるのも理解を深めるのに役立ちます。  
  
このようにして、距離と時間の関係を理解し、計算することができます。何か質問があれば、いつでも聞いてくださいね。