

## ChatGPT を活用した学習支援システムの開発と 医療情報分野の学習における有効性の評価

大井悠成<sup>\*1</sup> 虫明昌一<sup>\*1</sup>

### 要 約

大規模言語モデル (Large Language Models : LLM) の登場により, 多くの分野で大きな革新がもたらされている。その代表格である ChatGPT は OpenAI が開発した自然言語処理モデルであり, 対話型のチャットサービスを提供している。幅広い知識と詳細な回答能力から, 医療分野の資格試験対策などの学習における活用が期待されている。しかし, 医療分野の学習において, ChatGPT の回答の精度や活用法が不透明なため, 不正確な情報や誤った返信が発生する確率の検証や, 正しい回答を得る確率を向上させるための質問文の模索, 医療分野の資格試験対策などの学習における活用法を検討する取り組みが求められる。この問題を解決するため, ChatGPT を活用した学習支援システムを開発し, 今回は特に, 医療情報分野の学習における有効性を検証することを目的とした。今回開発したシステムでは, ChatGPT API を利用して学生の問い合わせや要望に対して迅速な対話形式の回答を提供する機能, さらに分野ごとに回答内容を保存する辞書作成機能を実装した。このシステムは医療情報技師能力検定試験を受験する学生を対象に使用し, 評価を行った。その結果, 全体的に高い評価を受けたが, 回答の長さや正確性に関する課題が浮き彫りになった。ChatGPT (GPT-3.5) の回答の正確性や適切性は61%であり, 誤回答の約7割は学生の質問の仕方に, 約3割は ChatGPT のデータセットの問題に起因していた。この結果から, ChatGPT を医療情報分野の学習に活用する際には, ChatGPT における教育やユーザーガイドラインの作成, 医療情報学用データセットを整備する仕組みの構築が重要であることが示唆された。

### 1. 緒言

近年, 急速な技術の進化により, 人工知能 (Artificial Intelligence : AI) は様々な分野で新たな可能性を切り開いている。また, 大規模言語モデル (Large Language Models : LLM) の登場により, 多くの分野で大きな革新がもたらされている。その代表格である ChatGPT<sup>1)</sup> は, OpenAI が開発した自然言語処理モデルで, 対話型のチャットサービスを提供している。ユーザーはプロンプトと呼ばれる入力領域に文字を入力し, ChatGPT は確率的に適切な回答を生成して返信する。文法や意味の一貫性を保ちながら, 自然な形で応答を行い, 過去のプロンプトも考慮して返答を生成する。利用者はユーザー

登録を行い, 無料版では GPT-3.5 を使用でき, 有料版の ChatGPT Plus では GPT-4.0 を利用可能である (2024年3月時点)。ChatGPT は多様な利用が可能であり, 幅広い知識と詳細な回答能力から, 様々な教育手法として活用されている<sup>2-4)</sup>。そのため, 医療分野の資格試験対策などの学習に対する活用の期待も高まっている。

しかし, 医療分野の学習において ChatGPT を活用する際の課題は, 誤情報への懸念である。本研究では, 医療情報技師能力検定試験を対象に, ChatGPT を活用した学習支援システムの有効性を検証する。文部科学省<sup>5)</sup> も生成 AI の適切な利活用の重要性を指摘しており, 教育現場での事例共有と

<sup>\*1</sup> 川崎医療福祉大学 医療福祉マネジメント学部 医療情報学科  
(連絡先) 大井悠成 〒701-0193 倉敷市松島288 川崎医療福祉大学  
E-mail : ooi@mw.kawasaki-m.ac.jp

指針の整備が求められている。

よって、本研究は、ChatGPT を活用した学習支援システムを開発し、今回は特に、医療情報分野の学習における有効性について検証をすることを目的とし、筆者らが対策に取り組んでいる医療情報技師能力検定試験を対象とする。本検定試験は、試験科目が、医学・医療系、情報処理技術系、医療情報システム系の3系に分かれており、医療情報分野の知識を修得していることを証明する資格試験である。合格には3系それぞれにおいて基準点を満たす必要があるため、全国の合格率が40%にも満たない難易度の高い資格試験である<sup>6)</sup>。また、医療情報分野は急速に進化する分野であり、正確な情報の理解が資格試験の合格に直結する。本研究を通して、ChatGPT を活用した学習支援システムが医療情報分野の学習にどれほど有益であるかを明らかにすることで、医療分野の教育における ChatGPT 活用の新たな役割を果たす可能性を追求する。

## 2. 方法


ウェブアプリケーションやサーバーサイドのスク립ト言語として、広く利用されている言語の一つである PHP を用いて、ChatGPT を活用した学習支援システムを開発する。そこに ChatGPT API を組み込むことで、学生の問い合わせや要望に迅速に対話形式で ChatGPT が返答することができる機能の実装を可能とした。実装した機能の画面を図1に示す。また、ChatGPT が返答した内容を、分野

ごとにシステム内に保存できる辞書作成機能も追加した。回答を保存する画面を図2、保存した回答を確認する画面を図3に示す。これらの機能を総じて、AI 機能と呼ぶ。また、医療情報技師能力検定試験の過去問題を解く機能も搭載している。このシステムを A 大学の医療情報技師能力検定試験を受験する学生（18名）を対象に資格試験対策として使用し、有効性の評価を行う。使用期間は2023年7月24日～8月20日とした。評価方法は、学生からのアンケート調査を通じて、AI 機能に対する評価の収集を行うこととした。また、データベースに格納している ChatGPT と学生とのやり取りのログを使用し、正確性や適切性を医療情報分野における2名の専門家がチェックすることで評価を行う。具体的には、専門家は ChatGPT の回答が医学的・情報学的根拠に基づき正確で、学生の質問に対して医療情報技師能力検定試験に関連した内容の回答を提供しているかどうかを基準に、ログを通じて1つ1つ確認し、評価を行う。ChatGPT のモデルは、Web 上で無料で使用できるため、学生が日々使用しているモデルである点、API 使用料が GPT-4.0 に比べて安価な点から GPT-3.5 で検証した。なお学生には事前に ChatGPT のメリット、デメリットを伝達し、学生と ChatGPT のやり取りのログを残したデータベースを逐一確認する体制をとった。誤回答を見つけた場合は、その都度対象学生に対して Microsoft Teams のチャット機能を通じて誤回答を訂正することで、学習に支障をきたさないように配慮した。

AIに質問

質問を入力してね:

セカンドオピニオンとは

 回答を考えています..

Q19: RAIDとは

A: RAID (Redundant Array of Independent Disks) は、データストレージの方式であり、複数の物理的なディスクドライブを組み合わせて1つの論理的なドライブとして使用する技術です。RAIDは、データの冗長性、パフォーマンスの向上、信頼性の向上などの利点を提供します。

RAIDでは、複数のディスクドライブをストライピング、ミラーリング、パリティ、またはこれらの組み合わせなどの方式で組み合わせて使用します。

- ・ストライピング (ストライプRAID) : データを断片化し、複数のドライブに分散して保存することで、データのパフォーマンスを向上させます。
- ・ミラーリング (ミラーRAID) : データを同じ内容の複数のドライブにコピーすることで、データの冗長性と信頼性を向上させます。
- ・パリティ (パリティRAID) : データとパリティデータ (エラー復元用の情報) を組み合わせて保存し、データの完全性を確保する仕組みです。

RAIDは、個々のディスクドライブの障害に対しても耐性を持たせることができます。たとえば、1つのドライブが障害を起こしても、RAIDは他のドライブからデータを復元できる場合があります。

RAIDは主にビジネス向けのデータストレージシステムやサーバーで使用されますが、個人ユーザーにも利点があります。

保存

評価

良い

普通

あまり

Q18: コホート研究とは

A: コホート研究とは、一定の要素 (コホート) を共有する特定のグループを対象に、時間の経過とともにその要素の変化や影響を追跡していく研究方法です。コホートは、通常は特定の時点で共通の特性を持つ人々のグループを指します。研究は、コホートが特定の要素に曝された後の健康状態や行動などの変化を追跡することで、要素とその変化との関連性を明らかにします。コホート研究は主に長期間にわたる研究であり、要素の変化や影響を追跡するための大規模なデータ収集が必要です。このような研究は、特定の要素が健康や疾病などに与える影響を評価したり、特定の時間帯や地域で特定の要素が増加することによる影響を調べたりするために使用されます。

保存

評価

良い

普通

あまり

図1 学生の問い合わせや要望に対話形式でChatGPT が返答する機能の画面

### 3. 結果

医療情報分野の学習における ChatGPT を活用した学習支援システムの有効性を、アンケート調査による評価と ChatGPT の回答の正確性や適切性の評価に分けて報告する。

#### 3.1 アンケート調査による評価

ChatGPT を活用した学習支援システムが医療情報系の資格試験対策に効果的であるかどうかを評価するために、5件法で回答する質問を設定した。その5つの項目の結果を図4に示す。いずれも5件法の回答の割合を円グラフで示している。また、自由記述で回答する質問を設定した4つの項目の結果を図5に示す。

図4より、「AI 機能は使いやすかったですか」「回答の速度は満足いくものでしたか」という質問に対して、すべての学生が「そう思う」「ややそう思う」と回答した。

と回答した。「回答や説明は理解しやすかったですか」という質問に対して、約9割の学生が「そう思う」「ややそう思う」と回答した。「回答を辞書に保存する機能は満足いくものでしたか」「AI 機能は今後の医療情報系の資格試験対策にも使いたかったですか」という質問に対して、約8割の学生が「そう思う」「ややそう思う」と回答した。

図5より、「AI 機能はどういう場面で利用しましたか」という質問に対して、学生は「その用語について、テキストなどで調べて理解できなかったとき」「インターネットで調べてみてもよく分からなかった単語を調べるとき」「気になった単語を手軽に調べたかったとき」などと回答した。「どのような回答の時に辞書に保存していましたか」という質問に対して、学生は「重要であると判断したとき」「理解しやすい内容が返ってきたとき」「説明が分かりや

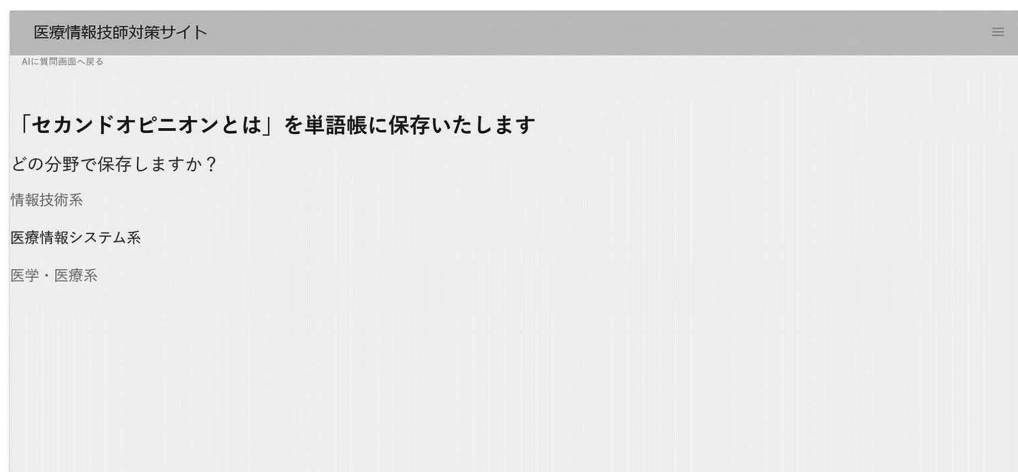


図2 回答を保存する画面（辞書作成機能）



図3 保存した回答を確認する画面（辞書作成機能）

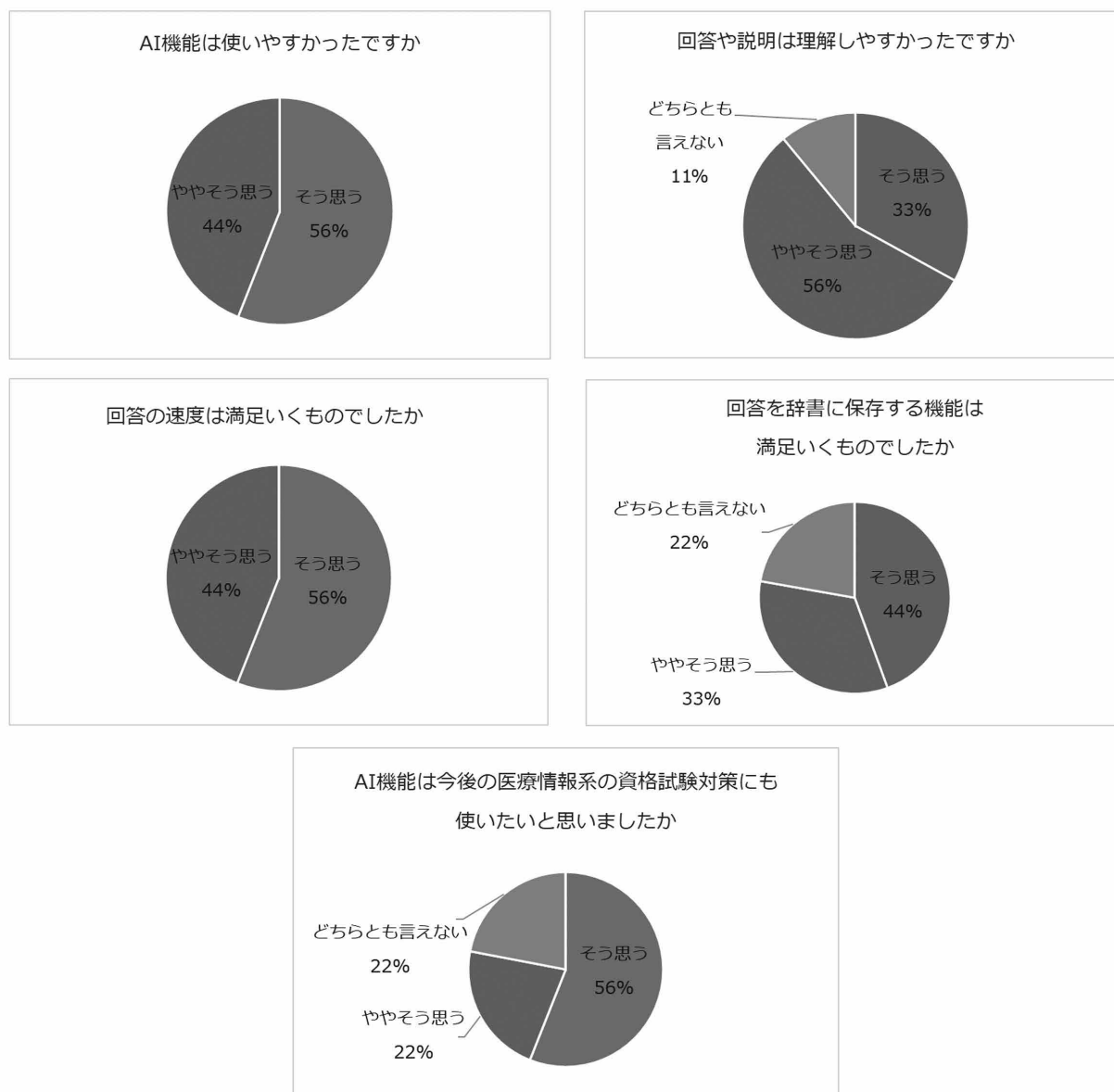


図4 AI機能に対するアンケート結果(5件法)(N=18)

すくて、見返したい回答が返ってきたとき」などと回答した。「AI機能があるおかげで、助かったことはありますか」という質問に対して、学生は「手軽にすぐに調べることができた」「教科書やインターネットで調べる手間が省けた」「教科書を使わず手軽に調べられてよかった」などと回答した。「AI機能で困ったことや改善点はありましたか」という質問に対して、学生は「文章が長くて、分かりづらいものがあつた」「中国語で表示される」「回答が正しくなさそうなものがあつた気がする」「難しい内容だったものがいくつかあつた」などと回答した。

### 3.2 ChatGPTの回答の正確性及び適切性の評価

ChatGPTと学生とのやり取りのログを、医療情

報分野の2名の専門家が確認し評価を行った。チェック対象を医療情報分野(医学・医療系、情報処理技術系、医療情報システム系の3系に関する質問)に限定し、ChatGPTの回答が医学的・情報学的根拠に基づき正確であるか、また学生の質問に対して医療情報技師能力検定試験に関連した内容の回答が提供されているかどうかを基準として評価を行った。評価した結果、全体の質問数147に対して正確性及び適切性が61%であった。また、質問内容の8割が分からない単語を問う質問であり、残りの2割がガイドラインの内容を問う質問であった。医療情報分野における分野ごとの質問数を表1に示す。表1より、医学研究に関する質問が最も多く寄せられていた。

<p>Q : AI 機能はどういう場面で利用しましたか</p> <p>A :</p> <ul style="list-style-type: none"><li>・ その用語について、テキストなどで調べて理解できなかったとき</li><li>・ インターネットで調べてみてもよく分からなかった単語を調べるとき</li><li>・ 気になった単語を手軽に調べたかったとき</li><li>・ 分からない単語があったとき</li></ul>
<p>Q : どういう回答の時に辞書に保存していましたか</p> <p>A :</p> <ul style="list-style-type: none"><li>・ 重要であると判断したとき</li><li>・ 回答を見て、覚えておいた方がいいなと思ったとき</li><li>・ 理解しやすい内容が返ってきたとき</li><li>・ 説明が分かりやすく、見返したい回答が返ってきたとき</li><li>・ 欲しい回答が返ってきたものは全て登録した</li></ul>
<p>Q : AI 機能があるおかげで、助かったことはありますか</p> <p>A :</p> <ul style="list-style-type: none"><li>・ 手軽にすぐに調べることができた</li><li>・ 教科書やインターネットで調べる手間が省けた</li><li>・ 教科書を使わず手軽に調べられてよかった</li><li>・ 調べても分からなかった単語の意味を理解することができた</li></ul>
<p>Q : AI 機能で困ったことや改善点はありましたか</p> <p>A :</p> <ul style="list-style-type: none"><li>・ 文章が長くて、分かりづらいものがあった</li><li>・ 中国語で表示される</li><li>・ 回答が正しくなさそうなものがあった気がする</li><li>・ 難しい内容だったものがいくつかあった</li></ul>

図5 AI 機能に対するアンケート結果（自由記述）(N=18)

誤回答を引き起こした原因を質問の仕方の問題とデータセットの問題に分け、それぞれの詳細の誤回答数と割合を示したものを表2に示す。質問の仕方を再構築して、質問したことで正しい回答が得られた質問を「質問の仕方の問題」、質問の仕方を再構築しても正しい回答が得られなかった質問を「データセットの問題」にしている。誤回答の約7割は学

生の質問の仕方に起因しており、約3割は ChatGPT のデータセットに問題があったことにより起因していた。質問の仕方の問題を細分化すると、質問の仕方により起因している質問の48%が医療分野以外でも使用される単語に対して、「医療」や「医学研究」などの文言を付与せずに質問しているため、学生が求める回答を得られていなかった。これらの質問に

対して、キーワードとして「医療」や「医学研究」を冒頭に加えて質問することで、意図に沿った回答が得られた。また、40%は漢字のみ、または英語のみの質問であったため、回答が中国語や英語で返されるケースが見られたが、「〇〇について教えてください」など日本語を加えることで日本語の正確な回答が得られた。13%は構文に不備があり、カタカナやひらがなが不自然に混ざっている、不要なスペースが含まれている、または疑問形になっていないなどの問題が含まれていたため、学生が求める回答を得られていなかった。これらの質問に対して、表記を統一し、構文を疑問形に変更するなど適切に修正することで、意図に沿った回答が得られた。データセットの問題を細分化すると、データセットの問題により起因している質問の39%は日本独自の取り組みに関する質問であり、33%はガイドラインの内容を問う質問であり、28%はその他の質問である。誤回答を起因した質問とその回答の例を項目ごとにまとめたものを図6に示す。図6の ChatGPT の回答は最初の部分だけを抜粋している。

#### 4. 考察

学生からのアンケート調査では、全体的に高い評価を示した。AI 機能は使いやすく、回答の速度や説明の理解が満足いくものであるとの回答が得られていた。また、自由記述回答から、学生は特に、インターネットや教科書で調べても分からなかった単語や手軽に単語を調べたい時に AI 機能を利用し、教科書やインターネットで単語を調べる手間が省けるといった意見が挙げられていた。これらから、ChatGPT を活用した機能によって、学生は迅速かつ容易に必要な情報を得ることができ、その手軽さや迅速な回答提供が学習支援システムのユーザーエクスペリエンス (User Experience : UX) 向上に寄与していると考えられる。また、質問内容の8割

が分からない単語を問う質問であったことから、学生は主に分からない単語を調べるために使用していたことが示唆される。

ChatGPT が回答したものをシステム内に保存することができる辞書作成機能は約8割の学生が高評価を示しており、自由記述回答から、学生は特に、ChatGPT の回答が重要であると判断したときや、理解が容易であり、将来的に見返したい回答が返ってきたときに辞書作成機能を使用していることが分かった。ChatGPT を活用した学習支援システムに辞書作成機能を追加することは、学習支援システムの更なる有益性を向上させる方向に寄与することが示唆される。ChatGPT の回答が手軽に保存できることで、学生は時間をかけずに効果的な情報の蓄積が可能となり、学習の時間効率が向上し、効果的な学習環境が整備される。資格試験対策なら、分野別に保存できるようにすることで、さらに利便性が高

表1 医療情報分野における分野ごとの質問数

分野	質問数
医学研究	44
医療プロセス	15
医療制度	14
臨床医学	13
医療システム	10
看護	10
検査	10
情報技術	10
診療記録	7
医療経営	7
その他	7
合計	147

表2 誤回答を引き起こした原因とその詳細別の誤回答数と割合

原因	原因の詳細	誤回答数	全体で占める割合	原因別で占める割合
質問の仕方の問題	「医療」や「医学研究」などの文言が付与されてない質問	19	33%	48%
	漢字のみ、英語のみの質問	16	28%	40%
	構文がおかしい質問	5	9%	13%
データセットの問題	日本独自の取り組みに関する質問	7	12%	39%
	ガイドラインの内容に関する質問	6	10%	33%
	その他の質問	5	9%	28%

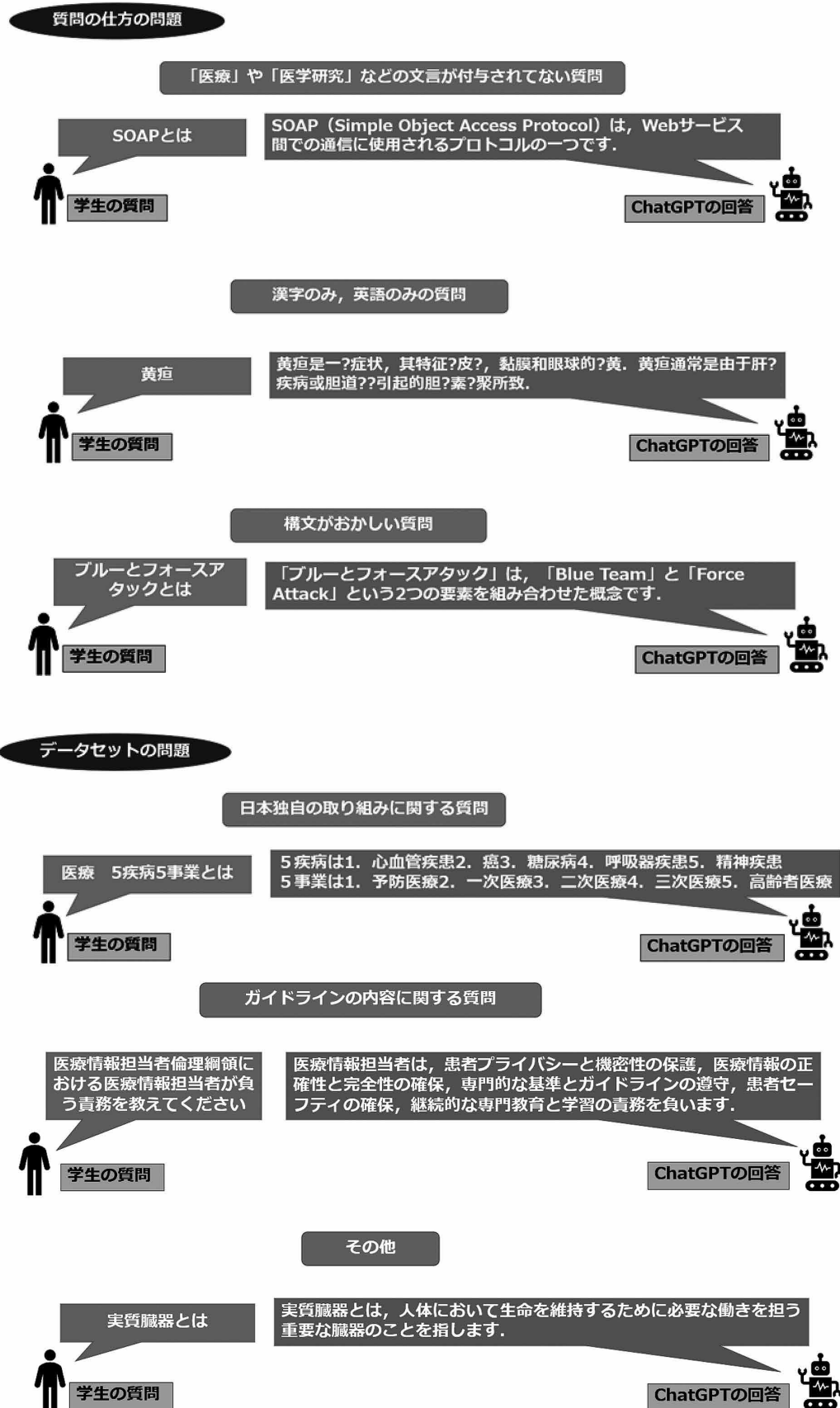


図6 誤回答を起因した質問とその回答の例を項目ごとにまとめた図

いものになると考えられる。

一方で、「AI 機能で困ったことや改善点はありませんか」という質問に対して、回答の長さや、理解しにくい回答が返ってくるといった課題を指摘している。これらは、ChatGPT を活用した学習支援システムを教育現場で活用するにあたって、改善点として考慮すべきポイントである。回答の長さや理解しにくい回答が返ってくるといった課題に対しては、ChatGPT が生成する回答のトークンの最大数を設定できるため、最大数を短くするといったプログラムが必要であると考えられる。また、「簡潔に分かりやすく答えてください」といった文言を質問文の最後に自動付与するようにプログラムで設定することで、より効果的になると考えられる。

また、回答の正確性への疑念といった課題を指摘しており、ChatGPT と学生とのやり取りのログを医療情報分野の専門家が確認し、評価を行った結果が示す通り、医療情報分野の学習における ChatGPT の回答には一定の改善の余地があることが示唆された。誤回答を引き起こした原因を分析すると、誤回答の約7割が学生の質問の仕方に起因し、約3割は ChatGPT のデータセットが問題で起因していることが明らかになった。

質問の仕方の問題は、医療分野以外でも使用される単語に対して、「医療」や「医学研究」などの文言を付与せずに質問しているケース、漢字のみ、または英語のみで質問しているケースが回答の不正確性の主な原因であった。よって、「医療」や「医学研究」や「〇〇について教えてください」などの特定のキーワードを自動付与するようにプログラムで設定する必要があると示唆された。また、寺岡<sup>7)</sup>は ChatGPT 活用に対するガイドラインを策定することで、ChatGPT が有効な教育支援ツールになると挙げており、朴ら<sup>8)</sup>は韓国で ChatGPT を、大学生を対象に学習に活用させた時の指摘内容<sup>9)</sup>から、質問またはプロンプトの提示能力が非常に重要であり、質問を生成する訓練を、生徒が ChatGPT を活用する前に行わなければならないこと、教員は倫理的ガイドラインを開発して共有し、過度に ChatGPT に依存しないように指導しなければならないと指摘している。これらから、ChatGPT における教育、質問の仕方の訓練や誤回答を起因する質問をもとに作成したユーザーガイドラインの提供をすることで、回答の正確性や適切性がさらに向上すると示唆された。ユーザーガイドラインの例としては、「医療」「医学」「〇〇について教えてください」といったキーワードを加えて質問内容を明確にすること、また「簡潔に」「具体的に」といった

文言を付け加えて回答の形式を指定することが挙げられる。さらに、不要なスペースや不自然なカタカナ・ひらがなをなくし、標準的な日本語表現を用いることで、質問内容が正確に伝わるような工夫も提案する。また、松田<sup>10)</sup>は ChatGPT の問題点として、「入力するデータが学習に使用され、外部に漏れる危険性」を挙げており、ネットリテラシー、情報の取り扱いなどの教育も重要である。そのため、ChatGPT を教育現場に活用するには、文部科学省の「初等中等教育段階における生成 AI の利用に関する暫定的なガイドライン」<sup>11)</sup>が示す通り、教員側がメリット、デメリットを把握するなどの一定の AI リテラシーが必要であると考えられる。

データセットの問題は、特に日本独自の取り組みに関する質問やガイドラインの内容に関する質問に対して、ChatGPT が正確な回答を提供できていないことが多かった。ChatGPT、特に GPT-3.5 のモデルを医療情報分野の学習に利用するには、その分野に特化したデータセットの整備が必要であると考えられる。特に日本独自の取り組みや医療制度やガイドラインに関する知識が GPT-3.5 のモデルの ChatGPT のデータセットに不足しており、英語圏以外の国の情報が特に不足している可能性が示唆された。データセットの問題であった質問について、モデルを GPT-4.0 に変更して検証した結果、ガイドラインの内容に関する質問を除いた質問で正確な回答が生成されていた。日本の医師国家試験に限定した研究でも、GPT-4.0 は過去六年分の問題で合格水準に到達したという報告<sup>12)</sup>の通り、GPT-4.0の方が回答の精度は優れており、GPT-3.5 は未整備な部分が多かったため、医療情報分野の学習においても、GPT-4.0の方が適切だと言える。しかし、ガイドラインの詳細な内容に関する質問で回答が正確ではなかった点、2024年3月現在、ChatGPT のデータセットは2022年1月<sup>1)</sup>までのデータしか持っていない点、また、松田<sup>10)</sup>は ChatGPT の問題点として、「生成される文章の正確性が保証されていない」「データにバイアスが含まれる」を挙げている点から、ChatGPT のデータセットに医療情報に関するガイドラインのデータや2022年以降の最新のデータを中心に医療情報学用のデータセットをリアルタイムで整備する仕組みを構築するといった対策が最適な解決策だと考えられる。どのモデルでも、ChatGPT を医療情報分野の学習に利用するにあたって、データセットの整備は必要不可欠である。専門分野の言語資源を教師データとして準備し、ファインチューニング<sup>†1)</sup>する研究<sup>13-14)</sup>は実施されており、専門知識をファインチューニングすることの



有用性が証明されている。医療情報学用のデータセットを ChatGPT のモデルにファインチューニングすることで、ChatGPT がより適切かつ正確な回答を提供できるようになると考えられる。そうなることで、学生が医療情報分野の学習において、より正確な知識を得ることができ、ChatGPT を活用した学習支援システムの質の向上が期待される。しかし、医療情報学用のデータセットを ChatGPT のモデルにファインチューニングするためには、医療情報分野に精通した専門家の協力が不可欠であり、これは適切なデータセットの整備という重要な課題を示している。さらに、医療情報分野は常に更新されており、最新の情報を反映するためにはデータセットを定期的に更新する仕組みが必要となる。これらの課題を考慮すると、多大な時間、労力そしてコストがかかることは避けられない。そのため、これらの課題を解決するため、大規模なプロジェクトとして計画し、適切なリソースを確保する必要があると考えられ、現時点において実現性は低いと言わざるを得ない。

また、仮にデータセットの整備やファインチューニングが実現した場合においても、ハルシネーション<sup>†2)</sup>のリスクを完全に排除することは困難である。ゆえに、ハルシネーションが発生する可能性があることを学生に認識させることが重要であり、ファクトチェックをするなどの対策が必要である。例えば、ファクトチェックの簡易な方法として、質問をそのままキーワードとして Google で検索し、複数の信頼性のあるソースを照合する仕組みを提供することが現実的である。この方法により、学生は自動化された検索結果を基に情報の正確性を判断しやすくなると考えられる。さらに、質問に対する回答の信頼性を自動的に評価し、参考文献や追加情報を提示する機能を追加することも、より高度なファクトチェック対策として有効である。この際、信頼に足る情報ソースの選定は重要であり、そのためのアルゴリズムの設計が鍵となることが考えられる。

とはいえ、「AI 機能は今後の医療情報系の資格試験対策にも使いたいと思いませんか」という質問に対して、約8割の学生が満足と今後の利用意向を示しており、ChatGPT が学生にとって有益で将来的な活用が期待されることが示唆された。今後、医療情報分野の知識獲得において有望な手段になると考えられ、学習支援システムと ChatGPT が連携することで、より効果的で主体的な学習環境を提供することが期待される。

## 5. 結語

本研究では、ChatGPT を活用した学習支援システムの開発とその有効性について検証を行った。その結果、医療情報分野の学習における ChatGPT の有効性が示され、学生からは高い評価を得ることができた。しかし、回答の長さや理解しにくさ、正確性への疑念といった課題も明らかになった。これらの課題に対する改善策として、ChatGPT における教育、ネットリテラシー、情報の取り扱いなどの教育はもちろん、質問の仕方の訓練や ChatGPT 活用におけるユーザーガイドラインの作成が必要であることが示唆された。また、医療情報分野におけるデータセットはまだ未整備な部分があるため、医療情報学用データセットといった、専門的なデータセットを整備する仕組みの構築が必要であることが示唆された。また、ハルシネーション対策として、ChatGPT を活用した学習支援システムを使用する際には、学生に ChatGPT の回答を鵜呑みにせず、必要に応じて他の情報源と照らし合わせるなど、情報の取り扱いや質問の仕方に注意を払うよう指導することを徹底していきたい。加えて、学生が簡易にファクトチェックを行える仕組みを実装し、回答の信頼性を評価するためのツールの提供や参考文献や追加情報を提示することで、学生へのハルシネーションの啓蒙を図ることも検討していきたい。

今後の展望として、医療情報学用のデータセットの整備や、誤回答を避けるためのプログラムの修正によって、学習支援システムの AI 機能の質がさらに向上していくことを期待したい。また、学生からのフィードバックを反映し、システムの使いやすさや機能性を向上させることも重要である。本研究は、一部の学生から得られたデータを基に評価を行ったが、参加者の人数が少ないため、より多くの学生からのデータを集め、結果の信頼性を高める必要がある。そのため、今後も、使用人数を増やし、ChatGPT を医療分野の資格試験対策などの学習に効果的に組み込む方法を模索し、機能向上を行っていく予定である。集まったデータをもとに、ユーザーガイドラインの作成、ChatGPT の使い方による学生への教育にも着手していく。

本研究は、ChatGPT を活用した学習支援システムの可能性を示す一例であり、今後もこの分野の研究が進むことで、より効果的で主体的な学習環境の提供が期待される。医療情報分野だけでなく、他の分野においても、ChatGPT のような AI ツールを活用した学習支援が広がることが期待される。

## 倫理的配慮

本研究の実施にあたっては、川崎医療福祉大学倫理委員会の承認を得た（承認番号：23-064）。

## 謝 辞

本研究は公益財団法人川崎医学・医療福祉学振興会の助成により実施した。

## 利益相反

著者らのいずれも開示すべき利益相反はありません。

## 注

- †1) ファインチューニングは、既存の学習済みモデルに独自のデータを追加して再学習させる技術であり、これによってモデルは新しい知識を取り入れ、カスタマイズされた性能を持つように調整される。竹内ら<sup>15)</sup>は LLM について、ファインチューニングと呼ばれる特定のタスクに特化した訓練を追加することで、テキスト分類・生成や感情分析、文章要約といったさまざまな自然言語処理タスクに適応できると言及している。
- †2) LLM におけるハルシネーションとは、事実に基づかない情報を生成する現象を指す。ハルシネーションは大きく2つに分類され、学習に用いたデータとは異なる事実を出力するケースと、学習に用いたデータには存在しない事を出力するケースに分けられる。ハルシネーションの原因は、データ、訓練、推論の三つの段階それぞれにあり、各段階において LLM がどのように知識を獲得し使用するかに深く関連しているとする報告もある<sup>16)</sup>。

## 文 献

- 1) Open AI : ChatGPT overview. <https://openai.com/chatgpt>, 2023. (2024.3.11確認)
- 2) 小原有以, 佐藤美唯, 倉光君郎 : KOGI—ChatGPT を Colab に統合したプログラミング演習支援—, 情報教育シンポジウム論文集, 2023, 141-148, 2023.
- 3) 尾関基行, 山本あすか : 遠隔グループディスカッションでの ChatGPT の利用に関する一検討. 日本教育工学会研究報告集, 1, 77-83, 2023.
- 4) 山本憲 : COVID-19パンデミックにおける医学教育の革新—VR 教材と医療面接チャットボットの開発と実装—, 医学教育, 54, 666-667, 2023.
- 5) 文部科学省 高等教育局 専門教育課 大学教育・入試課 : 大学・高専における生成 AI の教学面の取扱いについて (周知). [https://www.mext.go.jp/content/20230714-mxt\\_senmon01-000030762\\_1.pdf](https://www.mext.go.jp/content/20230714-mxt_senmon01-000030762_1.pdf), 2023. (2024.3.11確認)
- 6) 医療情報技師育成部会 : 医療情報技師の試験統計 (年度別認定数). <https://www.jami.jp/jadite/new/toukei/first/first-23total.html>, 2023. (2024.3.11確認)
- 7) 寺岡三左子 : ChatGPT® は有効な教育支援ツールになり得るか?—看護学における対話型 AI をめぐる議論の動向—. 医療看護研究, 20, 64-70, 2023.
- 8) 朴喆恩, 北村有人, 右田正夫 : 生成型人工知能が教育に及ぼす影響. 滋賀大学教育学部紀要 教育科学, 73, 219-224, 2023.
- 9) 한형중 : ChatGPT 의 교육적 활용을 위한 고려 요소 탐색에 관한 질적 연구. *문화기술융합학회논문지*, 9, 659-666, 2023.
- 10) 松田翔太 : ChatGPT を活用した情報教育の提案. 第一工科大学研究報告, 35, 37-40, 2022.
- 11) 文部科学省初等中等教育局 : 初等中等教育段階における生成 AI の利用に関する暫定的なガイドライン. [https://www.mext.go.jp/content/20230710-mxt\\_shuukyo02-000030823\\_003.pdf](https://www.mext.go.jp/content/20230710-mxt_shuukyo02-000030823_003.pdf), 2023. (2024.3.11確認)
- 12) Kasai J, Kasai Y, Sakaguchi K, Yamada Y and Radev D : Evaluating GPT-4 and ChatGPT on Japanese Medical Licensing Examinations. *arXiv*, 2303, 18027, 2023.
- 13) 八幡一紀, 桐生佳介, 片柳宏太, 毛利研, 関本和穂 : GPT-2を活用した質問応答タスクへの適応. 2022年度人工知能学会全国大会講演集, 36, 1-3, 2022.
- 14) 箱石健太, 一言正之, 菅田大輔 : 土木分野における事前学習モデル BERT による精度検証. 土木学会論文集, 79, 1-10, 2023.
- 15) 竹内広宜, 小川克也, 江頭俊亮, 小形真平, 山本修一郎 : エンタープライズアーキテクチャを用いた知識流通モデルによる生成系 AI の業務適用分析. 人工知能学会第二種研究会資料, 2023, KSN-033-03, 2023.

- 16) Huang L, Yu W, Ma W, Zhong W, Feng Z, Wang H, Chen Q, Peng W, Feng X, ...Liu T : A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *arXiv*, 2311, 05232, 2023.

(2024年11月14日受理)

## Development of a ChatGPT-Based Learning Support System and Evaluation of Its Effectiveness in Learning in the Medical Information Field

Yusei OI and Masakazu MUSHIAKI

(Accepted Nov. 14, 2024)

**Key words** : ChatGPT, medical informatics, learning support system, educational effectiveness, ChatGPT API

### Abstract

The emergence of Large Language Models (LLMs) has brought about major changes in many fields. ChatGPT, a natural language processing model developed by OpenAI, provides an interactive chat service. With its wide range of knowledge and detailed response capabilities, ChatGPT is expected to be used for learning in the medical field, for example, in preparation for certification exams. However, the accuracy of ChatGPT's answers and its utilization in learning in the medical field is unclear. Therefore, we developed a ChatGPT-based learning support system, and the purpose of this study was to verify its effectiveness in learning in the medical information field. The developed system implements a function to provide quick interactive responses to students' inquiries and requests using ChatGPT API, as well as a dictionary creation function that stores responses for each field. This system was used and evaluated for students taking the medical information Technician Proficiency Test. The results showed that the survey received high marks overall, but highlighted issues related to the length and accuracy of the responses. The accuracy and appropriateness of ChatGPT (GPT-3.5) responses were 61%, with approximately 70% of the incorrect responses due to the way students asked questions, and 30% due to problems with the ChatGPT data set. These results suggest the importance of education and user guidelines for ChatGPT users and the establishment of a system to maintain datasets for medical informatics when ChatGPT is used for learning in the medical information field.

Correspondence to : Yusei OI

Department of Health Informatics  
Faculty of Health and Welfare Services Administration  
Kawasaki University of Medical Welfare  
288 Matsushima, Kurashiki, 701-0193, Japan  
E-mail : [ooi@mw.kawasaki-m.ac.jp](mailto:ooi@mw.kawasaki-m.ac.jp)  
(Kawasaki Medical Welfare Journal Vol.34, No.2, 2025 349 – 359)