

◎総説

生成 AI を理学療法分野で活用するための基礎知識

Fundamental knowledge for utilizing generative AI in the field of physiotherapy

寺本優香¹⁾, 松下光範^{*2)}Yuka Teramoto¹⁾, Mitsunori Matsushita^{*2)}

要 旨

近年、生成 AI 技術、とりわけ大規模言語モデル (LLM) は様々な分野で活用されている。LLM は情報検索や文章生成、要約などの分野で高い性能を示し、多くの実務で利用されるようになってきた。しかし LLM の活用にはハルシネーションやバイアス、倫理的な問題といったリスクも伴う。とりわけ医療・ヘルスケア分野においては、誤情報の拡散や専門家による評価の必要性が指摘されており、慎重な運用が求められる。

本稿では、LLM の基本的な仕組みやプロンプトエンジニアリングなどの応用技術について具体例を交えて概説する。また LLM を用いた用語解説、症状の予後予測といった理学療法分野での LLM の活用可能性について簡単に検討するとともに、適用に際しての課題や注意点を整理する。これにより、理学療法の専門家が LLM を適切に活用するための基礎知識を提供し、今後の技術発展に向けた議論の礎とすることを目指す。

【キーワード】生成 AI、大規模言語モデル (LLM)、ハルシネーション、プロンプトエンジニアリング

1. はじめに

近年、大規模言語モデル (Large Language Model、以下 LLM と記す) あるいは生成 AI という単語を頻繁に耳にするようになった。2022 年に OpenAI により発表された ChatGPT は、一般に AI と聞いて想像されるような自然な対話を実現し、LLM ブームの火付け役となった。

LLM の特徴は、単なるテキスト生成にとどまらない高い情報要約能力や質問応答能力にある。これにより記事作成やプログラミングのサポート、情報提供など幅広い用途で注目されている。しかし LLM の活用には限界やリスクも存在する。とりわけ医療・ヘルスケアといっ

た人命にかかわる分野では、LLM の特性や問題点を把握し対策することが非常に重要である (Hicks et al., 2024)。本稿では、LLM の基礎、活用方法、注意点、さらに理学療法分野での応用について解説する。

2. LLM の定義とその歴史

本節では、LLM の前身となる言語モデル (Language Model) について解説する。

コンピュータの計算能力向上やインターネットの発達に伴い、大規模なデータ (ビッグデータ) の取り扱いが以前より容易になった。それに伴い、データの傾向を自動で取得し、予測や知見の獲得に役立つ機械学習が注目

1) 同志社大学大学院文化情報学研究所 Graduate School of Culture and Information Science, Doshisha University

2) 関西大学総合情報学部 Faculty of Informatics, Kansai University

公開日: 2025 年 3 月 31 日

【*責任著者】松下光範 関西大学 総合情報学部 〒 569-1095 大阪府高槻市霊仙寺町 2-1-1 A-321 m_mat@kansai-u.ac.jp

を集めはじめた。2000 年代以降発生し今日まで続く第 3 次人工知能ブームでは、人間の脳神経を模したニューラルネットワークと呼ばれるモデル構造が出現した。2024 年、ニューラルネットワークの基礎技術に関する功績で John J. Hopfield 氏と Geoffrey E. Hinton 氏にノーベル物理学賞が授与されたことは記憶に新しい。

ニューラルネットワークモデルは、テキスト・画像・音声といった様々な形式の情報を数値に変換したデータの入力を想定している。モデルの最終目的は、これらの数値データをネットワークを通してさらに変換していき、最終的に正解の状態を表現する数値の羅列を出力することである(図 1 参照)。理想的なモデルを目指し、ネットワークの内部構造を改良してつなぎなおす過程を学習と呼ぶ。学習は実際の正解状態を表現する数値の羅列と、それを目指した出力との乖離がある程度無くなるまで続けられる。こうして完成したモデルは未知のデータに対しても、課題を解決するうえで適切な数値への変換が可能となる。

現在ニューラルネットワークモデルのなかでも、オートエンコーダと呼ばれる仕組みが広く採用されている。オートエンコーダでは入力データを中間層まででより少ない数値列にまで圧縮し、出力層でもとの数値状態に戻

す。この学習方式により、人間が意味を解釈して理想の出力をモデルに教えることなく、ひたすら大量のデータを学習させることができる。オートエンコーダにおいて、中間層以前の部分はエンコーダといい、入力された数値データを圧縮して扱いやすいより少ない数値に変換する機能を持つ。一方中間層以降の部分はデコーダと呼ばれ、圧縮後の数値から元のデータ状態まで戻す機能を持つ。各部位はそれぞれ取り外し独立した機構として使用することができる。LLM の正体は人間の質問・回答の対を大量に学習させたオートエンコーダのデコーダ部分である。入力された質問がデコーダのエッジをたどることで、学習済みデータの傾向に最も近い回答込みのテキストに変換されているのである(Radford and Narasimhan, 2018)。

LLM という言葉の意味は本来「言語を扱う大規模なモデル」である。しかし ChatGPT の出現と同時に使用が広まった背景から、対話形式回答を出力し、ユーザと自然なやりとりを行うものを LLM と呼称することが多い。図 2 に LLM との対話実例を示す。プロンプトとはユーザから LLM に与えられる質問を意味する。プロンプトとは人間からの入力テキストを指す。

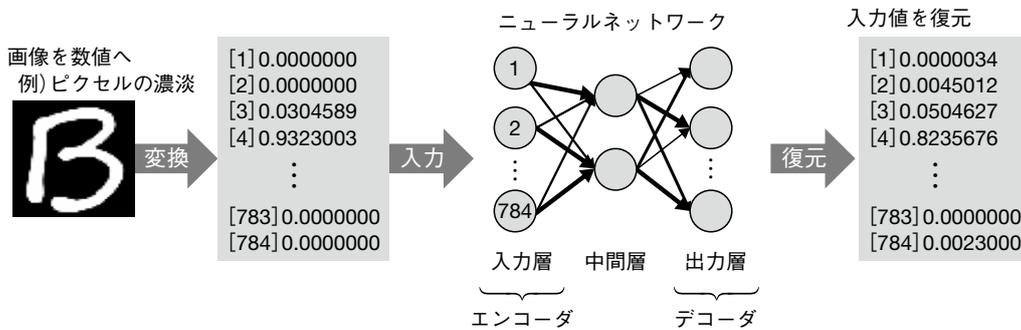


図 1 ニューラルネットワークの学習過程

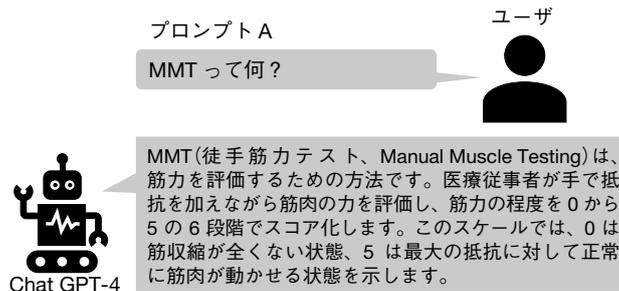


図 2 LLM との対話実例

3. 代表的な LLM のサービス

現在、様々な LLM が利用可能になっている。本節では、特に代表的な LLM サービスとして ChatGPT、Gemini、Claude を紹介する。

ChatGPT は米国の非営利法人 OpenAI により 2022 年 11 月に公開されたチャットボット型の WEB サービスである。2024 年 12 月 14 日現在、無償版では GPT-4、GPT-4o mini、GPT-4o、o1、o1-mini を制限付きで使用することが可能である。画像ファイルや音声等の入力に関しても一日の上限以内で実行可能になっている。

Gemini は、Google の Bard(2023 年 3 月公開)というサービスを前身としたチャットボット型の WEB サービスである。正式な公開は 2023 年 12 月であり、当時としてはまだ少なかった画像や音声での入出力に対応したマルチモーダルモデルであった。Gemini の強みは、Google の他のサービスと連携することで可能になる、個人に特化したアシスト機能である。具体的には G メールと連携してメールの下書きを自動で作成する、カレンダーの予定にアクセスする、Google Drive に保存している書類をもとに質問に回答する、といった作業が可能になっている。

Claude(クロード)は米国の AI スタートアップ企業「Anthropic」が開発した生成 AI モデルおよびそれを利

用したチャットボット型の WEB サービスで、ChatGPT に続くかたちで 2023 年 3 月に公開された。チャットボットの特徴として、多言語への対応・倫理面を重視したフィードバック・長文入力への対応といった機能を早期の段階から盛り込んだ点があげられる。

4. LLM におけるハルシネーション問題と RAG

LLM 以前の機械学習によって構築された言語モデルにおいて、人種や性別による差別など、学習データが含むバイアスを反映してしまう問題が指摘されていた(Sun et al., 2019)。LLM においてもなおその点に関しては注意が必要である。つまり、LLM を構築する際のデータに間違いや偏見が含まれていると、それが結果に悪影響を及ぼしてしまう可能性がある。一方、LLM 特有の問題点にハルシネーションが挙げられる。ハルシネーションとは虚構の情報をあたかも事実であるかのように生成する挙動で、前述したような高品質の製品版 LLM においても頻繁に確認される。図 4 は 2024 年 12 月 14 日現在、OpenAI の GPT-4 で確認された「ハリーポッター」の少々マニアックな知識に関するハルシネーションの例である。フィニアス・ナイジェリアス・ブラックという人物の子供を問われた LLM は、シリウス(二世)とファラクシアの 2 名がそうであると回答している。し

ハリーポッターに出てくるフィニアスナイジェリアスブラックの妻の名前は？

- ④ フィニアス・ナイジェリアス・ブラックの妻の名前はウルスラ・フリントです。彼女は純血の魔女で、フリント家の出身です。ウルスラは『ハリー・ポッター』シリーズで直接言及されることは少ないですが、ブラック家の家系図に名前が記されています。

フィニアス・ナイジェリアス・ブラックとウルスラ・フリントの子供の名前をすべて教えてください

- ④ フィニアス・ナイジェリアス・ブラックとウルスラ・フリントの子供たちは以下の通りです：
1. シリウス・ブラック (二世)
 2. ファラクシア・ブラック

これらの子供たちはブラック家の家系図に記されており、それぞれがブラック家の純血の血統を継承しています。

🔊 📄 📁 🔄 📄 📄

図 3 ハリーポッターの設定に関するハルシネーション例

走れメロスの登場人物をリストアップしてください

『走れメロス』に登場する主な人物は以下の通りです：

1. メロス—主人公で、正義感が強く、王に対しても真実を語る羊飼い。
2. セリヌンティウス—メロスの親友で、彫刻家。メロスが王に反逆の罪で捕らえられた際に、彼の妹と結婚式を挙げる予定だった。
3. 王 (ディオニス)—残酷で人を信じない王。メロスに対して試練を課す。
4. メロスの妹—名前は登場しないが、セリヌンティウスと結婚する予定の女性。

これらの人物が物語の中心をなしています。



図 4 走れメロスの設定に関するハルシネーション例

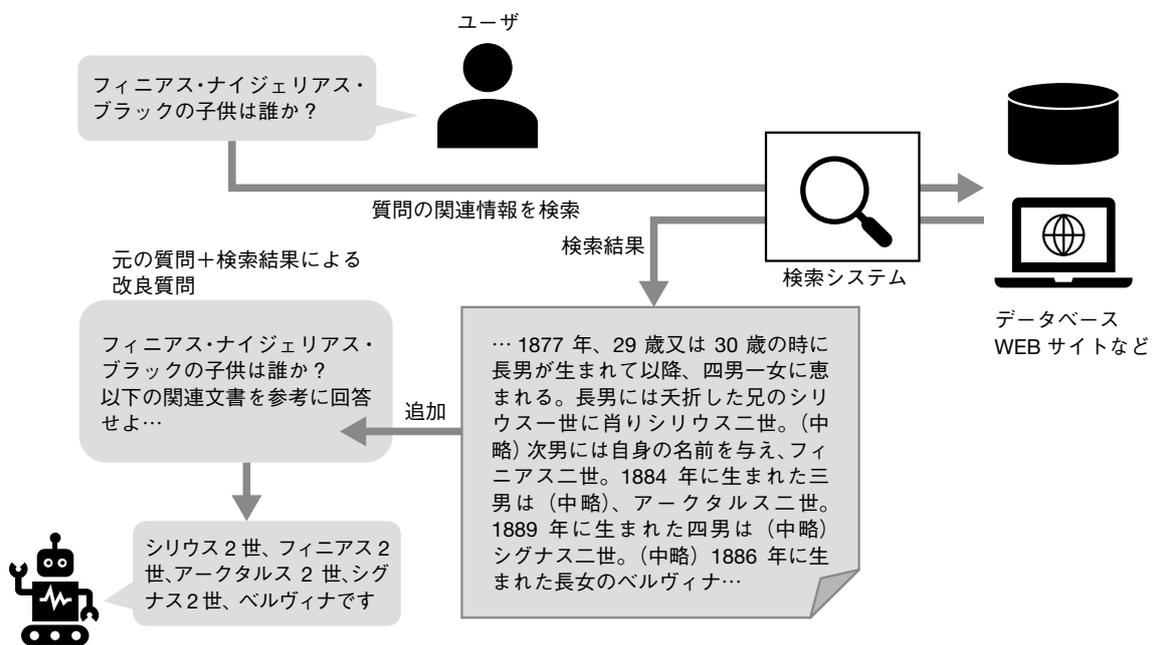


図 5 RAG の構成

しかし実際はシリウス 2 世、フィニアス 2 世、シグナス、ベルビナ、アクトゥルスの 5 名である。

より日本人にもなじみ深い例を追加で挙げる。図 5 は同じく 2024 年 12 月 14 日現在、OpenAI の GPT-4 に生成させた「走れメロス」の登場人物のリストである。セリヌンティウスとメロスの妹が結婚予定であるという原作には存在しない設定が追加されている。こうしたハルシネーションは主に以下の条件で引き起こされる。

- ・学習元のデータにとって未知の知識に関する生成である
- ・学習元のデータ自体に間違いが含まれている
- ・テキストを出力する過程で情報が欠落・挿入・結合により改変される

またハルシネーションにまで至らない場合でも、未知の知識を要するタスクにおいては LLM の性能が低くなることが明らかになっている。とりわけ専門の知識や信頼性を要する分野では、バイアスや偏見、ハルシネーション、学習データにおける知識の偏りによる低品質な回答に基づき誤った判断を下すことが重大なリスクにつながる。こうした諸問題への対策として近年 RAG (Retrieval-Augmented Generation) (Zhao et al., 2024) という手法が注目されている。RAG は、テキストを生成するプロセスに外部知識からの入力を組み合わせるアプローチである。WEB 検索の結果や参照先のデータベースがある程度信頼できる情報源であり、検索精度も十分高い場合、より確かな情報に基づく回答を得られる。一方で参照先

| 番号 | プロンプト原則 | 期待される効果 |
|----|---|------------------|
| 3 | タスクを単純なものに分割 | 複雑なタスクの実行 |
| 5 | 有効な説明依頼形式を利用 ・○○について簡単に説明してください ・11歳の子ども向けに説明してください ・その分野の初心者に説明してください | より深い理解につながる解説 |
| 7 | Few-shot prompting の導入 | より高精度な出力 |
| 14 | 必要な情報を得るまで質問を続けるよう指示 ・不明点があれば質問をしてください | 情報不足による不十分な回答の低減 |
| 16 | モデルの役割（例：教師、エンジニア）を想定 | ルールに沿った役割 |
| 19 | Chain-of-Thought (CoT) と Few-shot prompting の併用 | より高精度な出力 |
| 24 | 書出しを指定しての生成 ・ChatGPT はテキスト生成 AI であり に続けて説明して | 期待するフォーマットに沿った出力 |

図 6 プロンプト 26 原則の一例

データベースの設計、検索キーワードの選定、外部情報の信頼性評価といった、旧来の検索システムと地続きの検討課題も抱えている。

5. プロンプトエンジニアリングの基礎知識

ハルシネーションほど有害ではないものの、もう一つの LLM 使用時における注意点として、質問方法依存で回答性能が変化する点が挙げられる。質問のテキストに含まれる情報が回答に影響を与えるような挙動は、コンテキスト内学習 (in-context learning) と呼ばれる。コンテキストとは文脈を意味し、コンテキスト内学習の効果を狙い入力を工夫する技術はプロンプトエンジニアリング (Prompt Engineering) と総称される。プロンプトエンジニアリングの研究は近年盛んであり、「プロンプト 26 原則」(Bsharat et al., 2023) は特に有名である。図 6 にこの原則の一例を示す。先述の RAG も、入力を加工していることから一種のプロンプトエンジニアリングといえる。ここでは、初学者が押さえておくべきいくつかのプロンプトエンジニアリング技術を紹介する。

5-1 Few-shot prompting

Few-shot prompting では、数件程度の具体例 (インプットと期待されるアウトプット) を提示してモデルにタスク特性を理解させる。命題・グループの実例を入力し、その命題の真偽を回答させる場合を考える。

次の数の合計は偶数である # グループ : 1、3、6

解 : ○

次の数の合計は奇数である # グループ : 23、37

解 : ×

次の数の合計は偶数である # グループ : 390、890、47

解 : ×

次の数の合計は奇数である # グループ : 16、37、32

解 :

Few-shot prompting では上のように実際に回答まで行った例を複数入力する。これにこれまでの回答方法を参考にしつつ、最後の問題の回答を出力することが可能となる。Few-shot prompting を用いることで多くのタスクで正答率が向上し、出力形式も安定しやすくなる。一方で適切な例を事前に用意する必要がある。具体例無しで指示のみを出すプロンプトは、Zero-shot prompting と呼ばれる。

5-2 Chain-of-Thought (CoT)

CoT (Wei et al., 2024) は、プロンプトのテキスト内において問題解決に適していると思われる思考・行動のプロセスをテキストとして明示的に示したうえで、なぜその結論に至るかの根拠を示すようモデルに促す手法である。Few-shot プロンプトと組み合わせて用いることが多い。以下はその具体例である。

次の数の合計は偶数である # グループ : 1、3、6

解 : 合計は $1 + 3 + 6 = 10$ であるので○

次の数の合計は奇数である # グループ : 23、37

解 : 合計は $23 + 37 = 60$ であるので×

次の数の合計は偶数である # グループ : 390、890、47

解 : 合計は $390 + 890 + 47 = 1327$ であるので×

次の数の合計は奇数である # グループ : 16、37、32

解：

「合計は～であるので」のように一度合計を計算する箇所が追加されたプロセスに該当する。これ以外にも、単にプロンプトの最後に「段階的に思考せよ (Let's consider it step by step.)」という文言を追加するよりシンプルな手法も存在する。CoT は多段階推論を要求する複雑な数学問題や分析で特に有用である。また途中の推論を可視化することにより、人間が回答内容の妥当性を判断しやすくなるという可読性・説明可能性の面でのメリットも存在する。

6. 理学療法分野における LLM 応用事例

本稿の最終章ではこれまでの解説をふまえ、理学療法の分野における LLM の活用について探る。なおここで紹介する事例では LLM には ChatGPT-4 を使用している。

6-1 理学療法における用語の解説

プロンプトによって回答内容が異なるのは前述のとおりである。用語の分野に関する手がかりを含み、初心者向けの説明を指示されたプロンプト B (図 7) は、より詳細かつ平易で分かりやすい回答を得ることができている。一般に用語の定義説明においては比較的適切な回答

が出力される傾向にある。またその真偽の確認も比較的容易であるといえる。

6-2 理学療法における推論

図 8 は、用語定義からもう一步進んだ複雑な問いかけへの LLM の回答例である。プロンプト C の回答はおおよそ一般的な回答としては的を射ているように見える。しかし専門家からは、TKA の手術侵入路によって筋力低下の報告がだいぶ異なっている点を考慮すべきであるとの指摘を受けた出力である。プロンプト D は膝の屈曲角度と日常生活の関係について実用性の高い出力が得られた例である。汎用的な利用を想定した LLM サービスの場合、明確な答え・一般的な見解が広く共有された回答は得意であることが多い。一方で、一般性が低く個々の事例をより詳細に観察する必要があるケースは慎重に取り扱うべきである。

6-3 実利用に向けて

多くの LLM による WEB サービスでは、デフォルトでは入力内容を保存しモデルの学習や回答に役立てる設定が ON になっている場合が多い。とりわけ個人情報・職務機密を含む可能性があるデータ入力には注意が必要である。こうした設定は多くの場合ユーザ各自で設定変

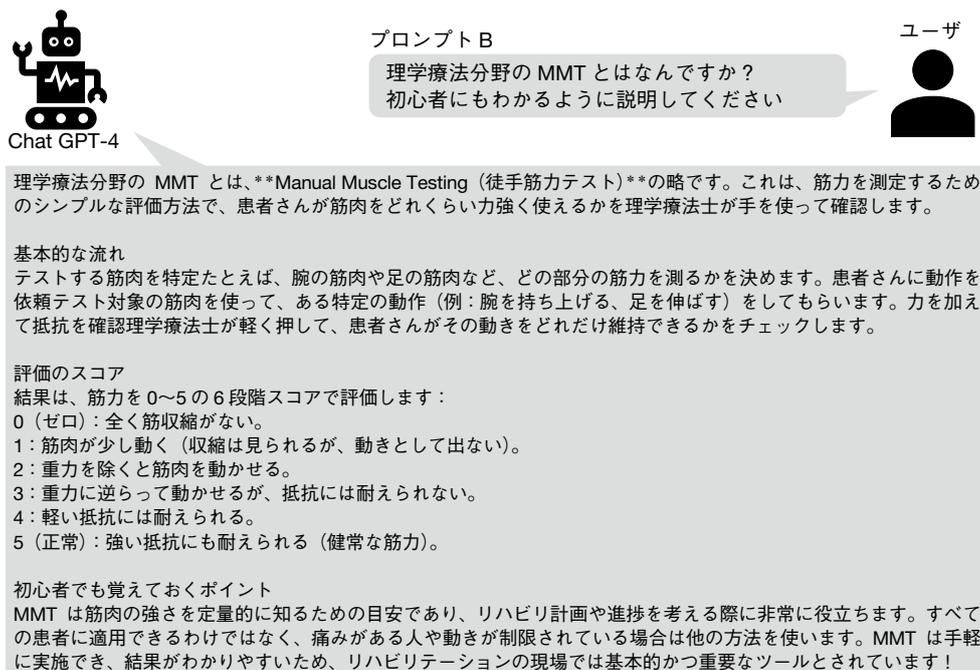


図 7 MMT の定義を出力するプロンプトの比較

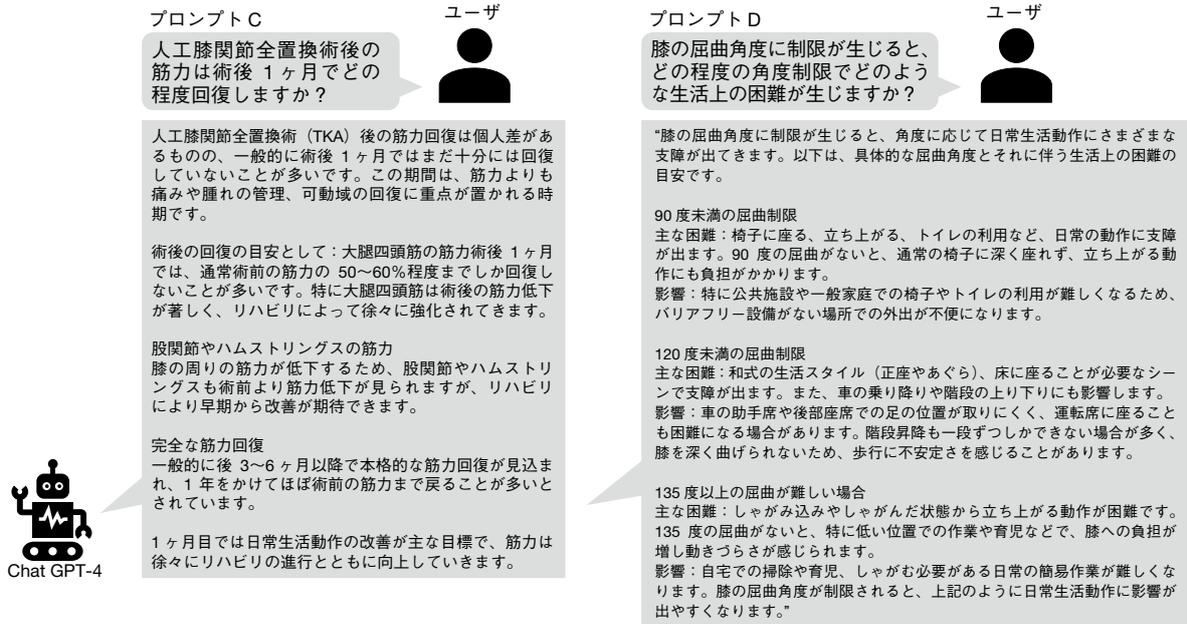


図 8 症状の予後予測および論理の組み立てタスクにおける出力例

更が可能である。しかしシステムの設定・規約確認に加え、入力データに関連する規約の確認もまた重要である。ここ数年、様々な専門分野での LLM 利用を取り扱った書籍が出版されており、理学療法の分野も例外ではない。現場の実問題に沿ったさらなる活用事例についてはそちらを参照されたい(松田雅弘他, 2024)。

7. おわりに

LLM は情報学分野において大きなパラダイムシフトをもたらし、現在盛んに研究が進められている。しかし当然ながら LLM の活用は情報学の分野のみに閉じているべきではない。今後あらゆる分野の専門家が LLM への理解を深め、適切に用いることで、幅広い分野の作業効率を劇的に向上させることができる。本稿を通じ、理学療法を専門とする方々が少しでも LLM に興味を持ち、様々な活用方法に思いを馳せてくだされば幸いです。

8. 謝辞

本稿を執筆するにあたり、同志社大学 文化情報学部メディア情報学研究室の波多野賢治教授、そして博士後期課程学生の木村優介さんのお二人からは有益なご意

見・ご指摘・アイデアを多数いただきました。両名にこの場を借りて深く御礼申し上げます。

参考文献

- Hicks, M.T. et al. (2024). "ChatGPT is bullshit." *Ethics and Information Technology*, 26(2).
- Radford, A. & Narasimhan, K. (2018). "Improving language understanding by generative pre-training." (Accessed: 19 December 2024).
- Scientist, D.K. (2020). "Retrieval Augmented Generation: Streamlining the creation of intelligent natural language processing models." (Accessed: 19 December 2024).
- Sun, T. et al. (2019). "Mitigating gender bias in natural language processing: Literature review." In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 1630-1640.
- Bsharat, S.M. et al. (2023). "Principled instructions are all you need for questioning LLaMA-1/2, GPT-3.5/4", <https://arxiv.org/abs/2312.16171>.
- Wei, J. et al. (2022). "Chain-of-thought prompting elicits reasoning in large language models." In *Proceedings of the 36th International Conference on Neural Information Processing Systems (NIPS '22)*, 24824-24837.
- 松田雅弘他 (2024)「セラピストのためのChatGPT活用ガイド 業務効率を最大化する賢いAIの使い方」三輪書店 ISBN-13 : 978-4895908382