

2024年3月11日

報道関係各位

GMOメディア株式会社

「日本語言語モデルの実践的評価」研究により IT分野での大規模言語モデル（LLM）の能力差を解明

GMOインターネットグループのGMOメディア株式会社（代表取締役社長：森輝幸 以下、GMOメディア）は、このたび「大規模言語モデルの日本語実践的評価：JGLUEとITパスポート試験を用いた比較分析」をテーマとした研究を実施しました。大規模言語モデル（LLM）が持つ推論能力（正確な解答を導き出す力）や問題解決能力の特徴を分析することで、LLMごとのIT分野における能力差を解明しました。

本研究論文は「2024年人工知能学会全国大会」（主催：一般社団法人人工知能学会）へ提出しています。また、詳細とその結果はコエテコAI教育研究所のページへ公開しています。

「大規模言語モデルの日本語実践的評価：JGLUEとITパスポート試験を用いた比較分析」
<https://college.coeteco.jp//ai-education-research-itpassport-results>

コエテコAI教育研究所

「日本語言語モデルの実践的評価」研究 IT分野での大規模言語モデル(LLM)の 能力差を解明

コエテコ by GMO

【本研究のポイント】

- ・現在公開されている大規模言語モデル（LLM）のIT分野における推論・問題解決能力を、ITパスポート試験を解答させた。
- ・結果から、IT分野の問題に対して一定程度論理的な解答を導き出すことが期待できることが判明した。
- ・最も解答の精度が高かったのは「ELYZA-japanese-llama-2-7b-instruct」で正答率は72.3%だった。
- ・LLMに補助的にITパスポート試験に関するヒント（プロンプト）を入力することで、ほとんどのLLMでヒントを与えない場合よりも出力（解答）の精度を上げることができた。

【研究概要】

■ 背景と目的

現在、各方面で ChatGPT をはじめとする大規模言語モデル (LLM) の評価が行われており、医療や法律といった特定の専門分野での有用性に関する評価も進められています。今回、IT 分野において LLM がどの程度の能力を持っているかを解明することを目的に研究を実施しました。

■ 研究内容

本研究では、IT に関する基本的な知識や応用能力を証明する国家試験である「IT パスポート試験」において、LLM が正解を導き出す力がどのくらいあるのかを指標としました。以下の LLM に「IT パスポート試験」の過去問を解答させることで、IT 分野の問題に対してどの程度正しい解答を導き出すことができるかを評価しました。

<研究対象の LLM>シリーズ名 (モデル名/開発元)

- GPT-3.5 (gpt-3.5-turbo-1106 / OpenAI)
- GPT-4 (gpt-4-11-6-preview / OpenAI)
- Japanese StableLM Alpha (Japanese Stable LM Instruct Alpha 7B v2 / Stability AI)
- Swallow (Swallow-7B-instruct-hf / 東京工業大学情報理工学院・国立研究開発法人産業技術総合研究所の研究チーム)
- Nekomata (nekomata-7b-instruction / rinna)
- ELYZA-japanese-Llama-2-7b (ELYZA-japanese-Llama-2-7b-instruct / ELYZA)

■ 研究方法

① : 汎用的な日本語問題と、専門的な日本語問題を使用した各 LLM の性能比較

入力される情報が特定分野へ特化していることが、LLM の解答の精度にどのように影響を与えるかを検証するため、LLM が一般常識をどの程度解答できるかを評価する「JGLUE 試験」^(※1) と「IT パスポート試験」の2つの試験に解答させ、正答率を検証しました。また、それぞれの結果を比較して分析を行いました。

(※1) JGLUE 試験：日本語 LLM の精度を評価する基本的な試験のひとつ。一般常識の質問に対する解答のスコアを基に LLM の精度を評価する。(詳細：<https://techblog.yahoo.co.jp/entry/2022122030379907/>)

② : プロンプトへのヒント挿入による LLM 推論能力 (正確な解答を導き出す力) の向上評価

LLM が解答を導き出す力をさらに検証するため、解答を導くために役立つヒントをプロンプトに追加し、ヒントが提供された場合に各 LLM が問題の正答率をどの程度向上するかを評価しました。

ヒントを含まないプロンプトで問題に解答させた①と、ヒントを含むプロンプトで問題に解答させた②の正答率を比較し、ヒントの有無が LLM の性能にどのような影響を及ぼすかの分析を行いました。

■ 研究結果

① : 汎用的な日本語問題と、専門的な日本語問題を使用した各 LLM の性能比較

「gpt-3.5-turbo-1106」「gpt-4-11-6-preview」「ELYZA-japanese-Llama-2-7b-instruct」は、約70%の精度で正しい解答を導き出すことができました。一方で、JGLUE 試験のような一般常識問題では高い正答率を出すことができる「Japanese Stable LM Instruct Alpha 7B v2」「nekomata-7b-instruction」は、IT 分野は苦手とする傾向が見られました。

② : プロンプトへのヒント挿入による LLM 推論能力 (正確な解答を導き出す力) の向上評価

IT 分野が苦手な LLM であっても、解答を補助するヒントを与えることで、ほとんどの LLM で解答の精度が向上することがわかりました。

<回答率一覧表>

LLM モデル名	JGLUE	IT パスポート試験 (ヒントなし)	IT パスポート試験 (ヒントあり)
gpt-3.5-turbo-1106	89.311	69.9	93.832
gpt-4-11-6-preview	95.42	70.53	82.99
Japanese Stable LM Instruct Alpha 7B v2	67.53	24.1	67.53
Swallow-7B-instruct-hf	46.91	31.1	62.5
nekomata-7b-instruction	81.68	58.8	86.16
ELYZA-japanese-Llama-2-7b- instruct	38.42	72.3	64.33

※LLM が獲得した点数割合 (%)

■ 結論

IT 分野での正確な解答を導き出す力は LLM のモデル間で差異があり、IT 分野に関して LLM に質問をする場合は、ある程度有効的に活用できる LLM (「gpt-3.5-turbo-1106」「gpt-4-11-6-preview」「ELYZA-japanese-Llama-2-7b-instruct」) と、活用に注意しなければならない LLM があることがわかりました。

【調査結果の活用方法と展望】

専門分野に特化するチューニングを施していない汎用的な LLM でも、一定の精度で IT 分野の正確な解答を期待できるため、IT パスポート試験の学習においても、誤りがある可能性を踏まえた上で、「分からない単語や概念について質問する」「練習問題を作成する」などの活用方法は、一定程度有効であると言えます。

本研究において、各 LLM に得意分野や不得意分野があることが判明したことは、価値があると考えています。今後も、こうした研究方法により他の資格試験を用いた LLM の評価分析を行っていく予定です。そして、それらの結果をもとに、資格試験をはじめとした学習に LLM を活用した授業の提案や対策講座の開発を促進してまいります。

GMO メディアは、今後も AI に関する研究を推進し、その結果を各事業へ生かす取り組みを積極的に行ってまいります。

【「コエテコ AI 教育研究所」について】

GMO メディアでは、2023 年 7 月に、AI の活用によって提供するサービスをさらに進化させるために AI チームを発足させました。AI チームは、AI を教育に適用させる研究を通じて、日本の教育産業に貢献することを目的に、プログラミング教育プラットフォームの「コエテコ byGMO」の知見を活かした研究を行う「コエテコ AI 教育研究所」を立ち上げました。本研究は「コエテコ AI 教育研究所」によるものです。

■ 論文執筆者

・羽中田 将 (GMO メディア株式会社 サービス開発部 シニアエンジニア)

現サービス開発部 AI チーム。2018 年入社。2020 年まで、技術推進室でレコメンドやデータ分析基盤の作成。その後、サービス開発部インフラチームの DBA(Database Administrator)で、MySQL の運用を担

当し、2023年よりAIチーム発足に伴い、社内データ活用やAIの推進を行う。

【GMOメディア株式会社について】（URL：<https://www.gmo.media/>）

GMOメディアは、創業以来インターネット上で自社開発・自社運営のサービス群であるメディア事業を中心に展開しています。現在はプログラミング教育ポータル「コエテコ」と美容医療の情報に特化した「キレイパス」の2事業を投資育成し、それぞれプログラミング教室や医療機関のDXを支援するサービスも展開しています。

さらに、ソリューション事業としてメディア運営で培ったノウハウを基にポイントサイトの構築を支援する「GMOリピータス」とアフィリエイトASP「AffiTown」など提携パートナーの収益化のサポートも行っています。

以上

【報道関係お問い合わせ先】

- | | |
|--|---|
| ●GMOメディア株式会社 | ●GMOインターネットグループ株式会社 |
| コーポレート部 広報担当 黒田 | グループコミュニケーション部 広報担当 青柳 |
| TEL：03-5456-2626 | TEL：03-5456-2695 |
| E-mail： pr@gmo.media | お問い合わせ： https://www.gmo.jp/contact/press-inquiries/ |

【GMOメディア株式会社】（URL：<https://www.gmo.media/>）

会社名	GMOメディア株式会社（東証グロース市場 証券コード：6180）
所在地	東京都渋谷区桜丘町26番1号 セルリアンタワー
代表者	代表取締役社長 森 輝幸
事業内容	■メディア事業 ■ソリューション事業
資本金	7億6,197万円

【GMOインターネットグループ株式会社】（URL：<https://www.gmo.jp/>）

会社名	GMOインターネットグループ株式会社（東証プライム市場 証券コード：9449）
所在地	東京都渋谷区桜丘町26番1号 セルリアンタワー
代表者	代表取締役グループ代表 熊谷 正寿
事業内容	■インターネットインフラ事業 ■インターネット広告・メディア事業 ■インターネット金融事業 ■暗号資産事業
資本金	50億円

※記載されている会社名・製品名は、各社の商標、もしくは登録商標です。

Copyright (C) 2024 GMO Media, Inc. All Rights Reserved.