



プレス通知資料（研究成果）

本件配布先：文部科学記者会、科学記者会、本町記者会

2023年9月19日

国立大学法人東京医科歯科大学

「ChatGPTが示す自己診断の信頼性の問題点とその改善法」 — ChatGPTに同じ症状の質問をしても、日によって違う診断を提示される？ —



【ポイント】

- 生成 AI が必ずしも正解を回答しないことが問題となっている昨今、整形外科疾患の自己診断における ChatGPT の信頼性を検証し、その診断の正答率および再現性の低さという問題点を提示しました。
- 質問の仕方次第で自己診断の信頼性が変化することを発見し、高い信頼性を得るための重要なポイントを明らかにしました。
- 自己診断ツールとして生成 AI が用いられるにあたっての安全性を向上させ、医療補助用にデザインされた新たな生成 AI システムの開発への大きな後押しとなることが期待できます。

東京医科歯科大学 統合イノベーション機構 オープンイノベーションセンター 医療デザイン部門の藤田浩二教授と大学院医歯学総合研究科 整形外科学の黒岩智之医師の研究グループは、代表的な生成 AI^{*1}である ChatGPT^{*2}を用いた自己診断において、5 日に渡り同じ質問を繰り返すことで、疾患によって正答率や再現性にばらつきがあり、特に広範囲に症状を生じる疾患において正答率および再現性ともに低いことを示しました。また質問の仕方次第でその正答率が変化することを見出し、どのような質問の仕方が ChatGPT による自己診断の信頼性を高めるかを提示しました。その研究成果は、国際科学誌 Journal of Medical Internet Research (ジャーナルオブメディカルインターネットリサーチ)に、2023年9月15日にオンライン版で発表されました。

【研究の背景】

近年、目覚ましい発展を遂げる AI の中でも、特に自然言語処理(NLP)^{*3}が大きな注目を集めています。ChatGPT は NLP を使用して自然な会話を作成できる最先端のチャットボットであり、爆発的な広がりを見せています。既に、生成 AI によるチャットボットを利用して病院受診前に自己診断を行う患者も増えてきており、今後増加の一途を辿ることが予想されます。しかしながら、自己診断における ChatGPT の正答率を評価した研究はいくつかあるものの、その再現性や受診勧奨の程度に関する研究はありませんでした。そこで我々は、ChatGPT による一般的な整形外科疾患の自己診断の正答率・再現性と、受診勧奨の程度を評価することを目標に研究を進めました。

【研究成果の概要】

本研究では、5人の研究者が、5つの整形外科疾患に関する質問を、5日間にわたり、全く同じ文面で ChatGPT (ver. 3.5) に繰り返し質問し、その回答結果を検証しました。疾患により正答率、再現性は異なり、最も低いものではわずか 4% の正答率で、再現性も「悪い」と評価されました。それにもかかわらず、回答内にて医療機関受診をしっかりと推奨していたものは全体の 13% 程度に留まりました。また、質問の仕方によって正答率が異なることを見出し、より好ましい質問の形式を提示しました。

【研究成果の意義】

生成 AI 技術が進化し爆発的な広がりを見せる昨今、痛みや不安を抱えた患者さんが生成 AI に自らの症状を相談して自己診断を行うことが予想されます。しかしながら、生成 AI が『ハルシネーション(幻覚)^{※4}』を生み出し使用者を混乱させうるという大きな問題も未だ残っています。そのような中で、本研究において代表的な生成 AI である ChatGPT の医療利用における問題点を浮き彫りにしたことで、その信頼性ならびに患者さんを害するリスクについての議論を生み出し、患者さんらへの啓発や、さらなるソフトウェア開発・進化の土台となることが期待されます。

今後、より多くの病態を対象に多岐にわたる質問形式を用いた研究を行うことで、病態ごとにより適切な質問方法を探していく予定です。また、ChatGPT 以外の生成 AI や新しいバージョンの ChatGPT を用いた研究を進めることで、得られる信頼性がどう変化するのも評価していく予定です。

【用語解説】

※¹ 生成 AI: Generative AI の訳語。従来の AI と異なり、非構造化学習データを元に、新しい情報・コンテンツを生み出すことができる AI のこと。

※² ChatGPT: OpenAI 社が 2022 年 11 月に公開した AI チャットボット。高度な自然言語処理技術により、人間のよう自然な回答をすることで耳目を集めた。代表的な生成 AI の 1 つ。

※³ 自然言語処理(NLP): Natural Language Processing の訳語。人間が日常的に使っている言語(自然言語)をコンピュータに処理・分析させる技術。AI と言語学の一分野。

※⁴ ハルシネーション(幻覚): 生成 AI の問題の 1 つ。問いかけに対して事実に基づかない情報を生成してしまう現象。AI がまるで「幻覚」を見ているかのように虚偽情報を出力することから名付けられた。誤学習や、データに無い内容に関して無理に回答を提示しようとすることで生じるとされる。

【論文情報】

掲載誌: Journal of Medical Internet Research

論文タイトル: The potential of ChatGPT as a self-diagnostic tool in common orthopedic diseases: An exploratory study

【研究者プロフィール】

藤田 浩二（フジタ コウジ） Fujita Koji

東京医科歯科大学 統合イノベーション機構

オープンイノベーションセンター 医療デザイン部門 教授

・研究領域

手外科、動作解析、疾患予測



黒岩 智之（クロイワ トモユキ） Kuroiwa Tomoyuki

東京医科歯科大学 大学院医歯学総合研究科

整形外科学分野 研究員

Mayo Clinic

Division of Orthopedic Surgery Research, Research Fellow

・研究領域

手外科、動作解析、超音波検査



【問い合わせ先】

<研究に関すること>

東京医科歯科大学大学院医歯学総合研究科

オープンイノベーションセンター 医療デザイン部門 藤田 浩二（フジタ コウジ）

E-mail: fujiorth@tmd.ac.jp

<報道に関すること>

東京医科歯科大学 総務部総務秘書課広報係

〒113-8510 東京都文京区湯島 1-5-45

TEL: 03-5803-5833 FAX: 03-5803-0272

E-mail: kouhou.adm@tmd.ac.jp