

Web 情報源の提示と意思決定

—救急要請判断の実証実験—*

伊藤愛美^a・小野木勇和^b・加藤秀明^c・木下和磨^d

木村仁咲^e・黒岩凜^f・小林愛^g・品田美咲^h

杉本七星ⁱ・須藤奏美^j・中田颯太^k・松本寛人^l

要約

本研究は、救急要請判断を支援する人工知能（AI）において、その判断根拠となる情報源を提示することが利用者の意思決定に与える影響を検証することを目的とする。AI の透明性を高めるための情報源の提示が、必ずしも適切な判断を促すとは限らず、逆に誤った判断を増加させる可能性について実験的に検討した。オンラインアンケート実験を実施し、314名の参加者を、AI の判断根拠（総務省消防庁のプロトコル URL）を提示する処置群と、提示しない統制群に無作為に割り付けた。参加者は、AI が提示する助言に基づき、複数のシナリオで救急車を呼ぶべきかを判断するタスクを行った。AI の助言には、正しいものと意図的に誤ったものが含まれている。分析の結果、特に「救急車を呼ぶべき」状況で AI が「呼ばないべき」と誤った助言をしたシナリオにおいて、情報源を提示された処置群の正答率は、統制群の正答率と比較して統計的に有意に低下した。

JEL 分類番号： D83, C90, I12

キーワード：生成 AI, 情報提供, 救急要請判断, デジタルヘルス

*なお、本論文に関して、開示すべき利益相反関連事項はない。

^a 創価大学経済学部経済学科 e2306222@soka-u.jp

^b 創価大学経済学部経済学科 e2306109@soka-u.jp

^c 創価大学経済学部経済学科 e2306319@soka-u.jp

^d 創価大学経済学部経済学科 e2306104@soka-u.jp

^e 創価大学経済学部経済学科 e2306428@soka-u.jp

^f 創価大学経済学部経済学科 e2306119@soka-u.jp

^g 創価大学経済学部経済学科 e2306127@soka-u.jp

^h 創価大学経済学部経済学科 e2306103@soka-u.jp

ⁱ 創価大学経済学部経済学科 e2306126@soka-u.jp

^j 創価大学経済学部経済学科 e2306426@soka-u.jp

^k 創価大学経済学部経済学科 e2306308@soka-u.jp

^l 創価大学経済学部経済学科 e2306212@soka-u.jp

1. はじめに

近年、深層学習に代表される人工知能（AI）技術は目覚ましい発展を遂げ、私たちの日常生活の隅々にまで浸透しつつある。特に、高度な専門知識が求められる医療・健康分野においても AI の活用が活発化しており、その診断能力は低くなく、現在もなお進歩し続けている（e.g. Takita et al., 2025）。

このような技術革新の流れの中で、市民が日常的に直面する医療判断の場面、とりわけ「救急車を呼ぶべきか」という緊急度判断の支援に AI を用いる試みに本研究は着目する。生命の危機に関わる場面において、生成 AI がもたらすハルシネーション（もっともらしいが事実と異なる情報）による誤った判断は、その費用（コスト）が大きい。ハルシネーションを軽減するためには、生成 AI が正しい専門知識を学習することが望ましい。また、ハルシネーションが起きているかを確認するために、その判断の基となった情報源を提示することも一つの手段として考えられる。しかしながら、本研究では、この情報源の提示が必ずしも誤った判断を軽減するとは限らず、かえって誤った判断を増加させることを明らかにした。

この新規性は、生成 AI を救急要請判断に用いることについて重要な示唆を与える。生成 AI の活用は、救急車の不適切な利用（不要不急の出動）の抑制と、逼迫する救急医療資源の適正化に貢献する可能性がある。一方で、解決すべき深刻な課題も存在する。最大の懸念は、AI の判断の信頼性と、それに対する人間の過信である。人は「根拠」を持ち出すことにより、異論や少数派意見を唱えやすく、社会的な非難を受けにくくなる（Bursztyn et al., 2023）。AI が症状の危険性を見誤り、救急車を呼ぶ必要がないと誤った助言をした場合、情報源の開示が、かえって利用者に情報源を確認せず利用することを促し、手遅れになるという最悪の事態も想定される。本研究は、この点を通して、AI が提示する情報を人間がどのように受け止め、最終的な意思決定に至るのかという、人間と AI のインタラクションのあり方について貢献するものである。

2. 実験デザイン

アンケートは Google Apps Script (GAS) を用いて作成し、Yahoo! クラウドソーシングにて回答者を募った。回答者を無作為に 2 グループに分けてアンケートを実施した。1つのグループは情報提供の際に参考元を提示したグループ（処置群）であり、もう1つは提示しないグループ（統制群）である。実験は 2025 年 8 月 30 日に成人の 10 代から 60 代までを対象として実施した。回答者 314 名のうち、男性は 231 名（73.6%）、女性は 83 名（26.4%）であった。また、10 代は 0.3%、20 代は 1.6%、30 代は 11.1%、40 代は 24.2%、50 代は 36.6%、60 代は 26.1%であった。

以下ではアンケートにて回答者が回答する順番に従い、各質問事項について説明する。

シナリオ「一昨日の夜にたくさんお酒を飲んだ45歳の男性は、それから2日以上強い吐き気・嘔吐が続いている。そのため飲食が全くだきていない。」この状況をAIに診断してもらいました。

呼ばない



参考：総務省消防庁
<https://www.fdma.go.jp/mission/enrichment/appropriate/appropriate002.html>

45歳男性の強い吐き気・嘔吐の症状について、「強い吐き気、または嘔吐が2日以上続いている」という点と、「飲食が全くだきていない」という点は確かに認められますが、重度の脱水や意識変化などの重篤な状態は現時点でないと判断されるため、**非緊急（低緊急・G）**と評価し、まずは自力受診を推奨します。（診断は必ずしも正しいとは限りません。重要な情報は確認するようにしてください。参考：消防庁「緊急度判定プロトコルver.3」
<https://www.fdma.go.jp/mission/enrichment/appropriate/appropriate002.html>）

☐ 救急車を呼ぶ ☐ 救急車を呼ばない

次へ進む

シナリオ「一昨日の夜にたくさんお酒を飲んだ45歳の男性は、それから2日以上強い吐き気・嘔吐が続いている。そのため飲食が全くだきていない。」この状況をAIに診断してもらいました。

呼ばない



45歳男性の強い吐き気・嘔吐の症状について、「強い吐き気、または嘔吐が2日以上続いている」という点と、「飲食が全くだきていない」という点は確かに認められますが、重度の脱水や意識変化などの重篤な状態は現時点でないと判断されるため、**非緊急（低緊急・G）**と評価し、まずは自力受診を推奨します。（診断は必ずしも正しいとは限りません。重要な情報は確認するようにしてください。）

☐ 救急車を呼ぶ ☐ 救急車を呼ばない

次へ進む

図1：処置群の表示画面（左）と統制群の表示画面（右）

- (i) 回答者は年代、性別の選択をする。
- (ii) 回答者はアンケートに関する注意書きを読み、それに対して同意する場合はチェックを入れ、「回答を始める」ボタンを押す。同意できない場合は回答終了となる。回答者は「回答を始める」ボタンを押した際に、無作為に2つのグループに割り当てられる。
- (iii) 回答者は、救急車を呼ぶべきかを判断するタスクを5回実施する。タスクの流れは以下の通りである。
- (iv) まず、症例シナリオを読み、終わり次第次のページに進む。
- (v) 図1に表示されている通り、症例シナリオをAIが判断し、結果が提示される。AIがこの判断を行う際、総務省消防庁の「緊急度判定プロトコル ver.3 119 番通報」を参照させる。この時、処置群はAIの判断の根拠（総務省消防庁の名前および緊急度判定プロトコル ver.3 119 番通報のURL）が提示され、統制群は判断の根拠が提示されない。その上で、救急車を呼ぶか、呼ばないかを選択する。なお、5回中初めの4回のAIの判断は、事前に症例シナリオをAIに判断させ、その結果を出力したものを使用している。
- (vi) 回答者は、(iv)～(v)を4回繰り返し行う。その際に表示される症例シナリオとAIの出力結果は表1にまとめられている。回答者は、この症例シナリオ1から症例シナリオ4までをランダムな順番で回答する。

表 1. 各シナリオとそれに対する生成 AI の出力

No.	正しい判断	AI の判断
1	救急車を呼ぶべき	救急車を呼ぶべき
2	救急車を呼ばないべき	救急車を呼ぶべき
3	救急車を呼ぶべき	救急車を呼ばないべき
4	救急車を呼ばないべき	救急車を呼ばないべき
5	救急車を呼ぶべき	プロンプトに依存

- (vii) 5 回目の症例シナリオでは、回答者は提示されたシナリオの状況を AI に診断してもらうための文章を、例を参考にして短文または箇条書き(10 文字以上)で作成し送信する。その内容を OpenAI の Application Programming Interface(API)を用いてリアルタイムで AI に判断を行ってもらい、その結果が提示される。その際に使用したモデルは GPT-4o mini である。これまでのタスクと同様に処置群には根拠となる URL が提示され、統制群には根拠となる URL が提示されない。その上で、救急車を呼ぶか、呼ばないかを選択する。
- (viii) 次に回答者は、デジタルヘルスツールである「Q 助」や「ユビー」の認知度や利用経験について、(利用したことがある・利用したことはないがどのようなものか知っている・聞いたことがある・知らなかった)の選択肢から回答を求めた。
- (ix) 最後に回答者は、AI への信頼度があるか(信頼している・信頼していない)から選択し、アンケートを終了する。

3. 仮説

設問 1~5 では統制群（情報源無し）と処置群（情報源有り）で AI の判断に対する信頼性が変わり、回答者の回答に対して影響を与えと考えられる。つまり、情報源がある場合は AI が正しい判断をしているかどうかに関わらず、それを信じてしまう。

また、救急車を呼ぶにあたる最終的な判断の責任を情報源に転嫁することができるため、自分の選択に対しての責任が軽くなり、AI の判断を採用するようになると考えられる。そのうえで、以下の仮説を考えた。

仮説 1

設問 1~4 でシナリオ 1(呼ぶべき時に AI が呼ぶべきと提示)とシナリオ 4(呼ばないべき時に AI が呼ばないべきと提示)では、情報源がある場合（処置群）の方が、情報源がない場合（統制群）より AI の信頼性が上がり、AI の判断のとおり回答者が正しい判断ができる。

仮説 2

シナリオ 2(呼ばないべき時に AI が呼ぶべきと提示)とシナリオ 3(呼ぶべき時に AI が呼ばないべきと提示)では、情報源がある場合(処置群)の方が、情報源がない場合(統制群)より AI の信頼性が上がり、AI の判断を鵜呑みにしてしまうことで、回答者が誤った判断をしてしまう。

4. 分析結果

処置群と統制群における正答率の平均値の差を用いて仮説 2 の検証を行った。図 2 は、設問 1~4 における各シナリオの回答者の正答率の平均値(縦軸)を示したものである。s=1~4 はそれぞれシナリオ 1~4 を示している(表 1 参照)、横軸のグループ A は統制群、グループ B は処置群を示している。

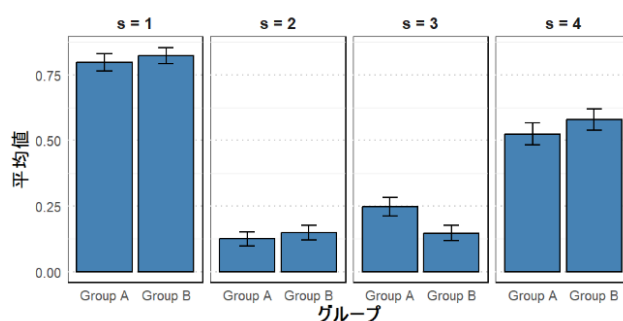


図 2 各シナリオにおける正答率の平均値の比較

シナリオ 3 に基づく回答結果を見ると、根拠なしの統制群の正答率は 0.247 であったのに対し、根拠ありの処置群は 0.147 であった。これらの正答率の差について、 t 値=2.256 が得られ、両側 5%水準で有意差が認められた。しかし、シナリオ 2 については有意差がなかった。したがって、仮説 2 のうち、シナリオ 3 に関しては支持されることが分かる。

5. 結論

本研究は、救急車を呼ぶべきかという緊急性の高い判断を支援する AI において、その判断根拠となる情報源を提示することが、利用者の意思決定にどのような影響を与えるかを実験的に検証した。特に、AI の助言が誤っている場合に、権威ある情報源(総務省消防庁)へのリンクを示すことが、かえって利用者を誤った判断に導く危険性について焦点を当てた。

分析の結果、仮説の一部が支持されることが明らかになった。とりわけ、救急車を呼ぶべき重大な状況であるにもかかわらず AI が「呼ばないべき」と誤った助言をしたシナリオ 3 において、情報源を提示された処置群は、提示されなかった統制群に比べて、有意に正答率が低下した。これは、利用者が AI の誤った助言を鵜呑みにし、本来であれば救急

車を呼ぶべき状況で「呼ばない」という危険な判断を下す割合が高まったことを示している。この結果は、情報源の提示がAIへの過信を生み、利用者が自ら情報源を確認したり、批判的に吟味したりすることなく、AIの助言に従ってしまう「自動化バイアス」を助長する可能性を示唆するものである。

一方で、AIの助言が正しい場合（シナリオ1および4）においては、情報源の有無が正答率に与える影響は限定的であった。これは、AIの助言が利用者の直感と一致する場合には、情報源の存在が判断を大きく左右する要因にはなりにくいことを示しているのかもしれない。

本研究がもたらす示唆は、医療分野におけるAIインタラクション設計の観点から極めて重要である。AIの透明性を高めるために判断根拠を示すというアプローチは、一見すると利用者の適切な判断を促すかのように思える。しかし、本研究の結果が示すように、特に生命に関わるような重大な局面においては、その情報源の権威性がかえって批判的思考を停止させ、AIへの盲目的な追従を促すという、いわば「諸刃の剣」となりうる。

今後の展望として、本研究の限界点も踏まえた更なる探求が求められる。本研究はオンラインアンケートによるものであり、実際の緊急時における心理的切迫感は再現できていない。また、情報源の提示方法についても、単なるリンク表示以外の方法（例えば、要約の提示や注意喚起メッセージの併記など）が利用者の判断に与える影響を比較検討する必要があるだろう。

結論として、救急要請判断支援のようなクリティカルな領域にAIを導入する際には、そのアルゴリズムの精度向上だけでなく、人間がAIの情報をいかに解釈し、最終的な意思決定を下すのかという認知プロセスを深く理解した上で、UI/UXを慎重に設計する必要がある。情報源の安易な提示が、時として悲劇的な結果を招きかねないことを、本研究は示している。

引用文献

- Leonardo Bursztyn, Georgy Egorov, Ingar Haaland, Aakaash Rao, Christopher Roth, Justifying Dissent, *The Quarterly Journal of Economics*, Volume 138, Issue 3, August 2023, Pages 1403–1451.
- Takita, H., Kabata, D., Walston, S.L. *et al.* A systematic review and meta-analysis of diagnostic performance comparison between generative AI and physicians. *npj Digit. Med.* **8**, 175 (2025)
- 総務省消防庁「緊急度判定プロトコル ver.3 119 番通報」
<https://www.fdma.go.jp/mission/enrichment/appropriate/appropriate002.html>