AI の意見が人間の倫理的判断に与える影響 ―大学生を対象とした実験的検討―*

内田柊也 a
鈴木颯和 b
武富真人 c
林杏香 d
柳田由芽 e
吉田蒼良 f

要約

近年、生成 AI の普及により、人が AI から助言を受ける場面が増えている。本研究は、AI の意見が人間の倫理的判断に与える影響を検討したものである。大学生 82 名を対象に、AI の回答を提示する条件と提示しない条件に分け、Greene ら（2001）の基準に基づく 6 種類の道徳的ジレンマ課題を実施した。分析の結果、いずれの課題においても AI 条件と非 AI 条件で判断の有意な差は見られず、AI の意見が参加者の判断に直接的な影響を与えることは確認されなかった。すなわち、参加者は AI の意見をそのまま受け入れるのではなく、参考程度に扱っていたと考えられる。このことは、教育・医療・法律などで AI を活用する際、最終的な判断を人間が担う重要性を示唆している。ただし、大学生のみを対象とした点や一部課題でわずかな変動が見られた点から、今後は多様な課題や対象者で検討する必要がある。

JEL 分類番号： C91, D91，O33
キーワード：AI，アルゴリズム忌避，アルゴリズム選好，道徳のジレンマ

---

* 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 成城大学社会イノベーション学部 23n2018c@u.seijo.ac.jp
b 成城大学社会イノベーション学部 23n2073b@u.seijo.ac.jp
c 成城大学社会イノベーション学部 23n2082h@u.seijo.ac.jp
d 成城大学社会イノベーション学部 23n2107h@u.seijo.ac.jp
e 成城大学社会イノベーション学部 23n2129m@u.seijo.ac.jp
f 成城大学社会イノベーション学部 23n2137a@u.seijo.ac.jp

1. イントロダクション

近年、生成 AI の普及により、人は専門家ではなく AI から助言やアドバイスを受ける場面が急速に増えている。AI からの助言を人がどのように受け止めるかという問題は、1950 年代（Meehl, 1954）から継続的に研究されてきた。近年の研究（Dietvorst et al., 2015; Logg et al., 2019）では、人間が AI や統計モデルを専門家よりも信頼しない「アルゴリズム忌避」と、逆に AI をより信頼する「アルゴリズム選好」の双方が確認されている。これらの知見は、人が AI から与えられる情報を人間の専門家からの情報とは異なる仕方で理解していることを示している。医療（Rao et al., 2023）、金融（Darwish et al., 2025）、教育（Mustafa et al., 2024）など幅広い分野で AI の導入が進む中で、アルゴリズム忌避や選好が実際の意思決定場面でどのように表れるのかを明らかにすることが重要な課題となっている。

ここで AI が提示する情報を人がどのように解釈し、自らの判断に組み込むのかという問題を考える上で興味深いのが道徳的ジレンマである。本研究で扱う「道徳的ジレンマ」とは、いずれの選択を取ってもある価値を犠牲にせざるを得ない状況を指す。典型的な例として、多数の人を救うために一人を犠牲にするかどうかといった問題が挙げられる。このようなジレンマは、人間が道徳的判断を下す過程を明らかにするために心理学や倫理学で広く利用されてきた。AI がこうした問題に対して提示する意見を人がどのように受け止めるかを検討することは、AI と人間の関係性を理解するうえで重要である。

以上を踏まえ、本研究の目的は AI の意見が人間の倫理的判断に与える影響を明らかにすることにある。大学生 81 名を対象に道徳的ジレンマ課題を提示し、参加者を二つの条件に分けて調査を行った。一方の条件では AI の回答を提示し、もう一方の条件では問題のみを提示した。両条件を比較することにより、AI の意見が人の判断に影響を与えるかどうか、またその影響の大きさを検討した。

2. 方法

82 名の大学生が参加し、40 名が AI 条件、41 名が AI 無し条件に割り当てられた。調査は Google フォームを用い、授業を利用して回答のフォームの URL を配布して行った。質問は Greene et al. (2001)で用いられた道徳のジレンマの中から，彼らの設けた Moral–Impersonal Dilemmas，Moral–Personal Dilemmas，Non-Moral Dilemmas の 3 つの基準に基づき，参加者が大学生であることを考慮して過度に残酷である，あるいは性的な話題を扱うものを避けて選択した 6 種類のジレンマ課題で構成した。具体的な課題を表 1 に示す。回答にあたっては、参加者はフォームの最初のページで性別・年齢に回答した後、

２ページ目のフォームである行為が適切かどうかを判断する課題であることを説明されたうえで、表１に示してある６種類のジレンマ課題を読み，登場人物の行為が「適切か否か」を評定した。AI 条件に割り当てられた参加者は表１に示された課題に回答し，AI なし条件では表１に示された文章から下線部に該当する箇所を削除された文章を読んで回答した。回答は Google フォーム上で行われ，所要時間は約 5 分程度であった。

表１　本研究で用いたジレンマ課題

会議の異動(non-moral)

午後 2 時に始まる会議に出席するため、ニューヨークからボストンまで移動する必要があります。電車かバスのどちらでも構いません。電車なら、どんな場合でも会議に間に合います。バスは会議の 1 時間前に到着する予定ですが、交通渋滞のため数時間遅れることもあります。会議の 1 時間前に余裕があれば良いのですが、遅刻は許されません。どうしようか悩んだ挙句，AI にバスに乗るべきか，電車に乗るべきかを聞いてみました．その結果，AI は電車に乗ることを勧めました．会議に遅れないようにするために、バスではなく電車に乗るのは適切か？

植物の移送(non-moral)

あなたは店から家まで約 5km の距離にある植物を運んでいる。 車のトランクは、植物の泥をキャッチするためにプラスチックを敷いており、車を汚さず植物を運べる。 1 回の運転で植物をすべて家に持ち帰ることもできるが、そのためにはトランクだけでは収まらないため後部座席にも植物を置く必要がある。AI に尋ねたところ、車の内装を台無しにしないために、植物を 2 回に分けて取りに行く方が適切であると答えました。後部座席に植物を置くと、数千ドルの費用がかかる高級革製内装を台無しにする。 車の内装を台無しにしないために、家に 2 回植物を取りに行くことは適切か？

ウィルス(moral-impersonal)

ウイルスが蔓延し、100 万人が死亡しています。あなたは家の実験室で 2 つの薬を開発しましたが、どちらがワクチンでどちらが致死薬かは不明です。唯一の識別方法は、2 人の患者にそれぞれの薬を投与し、結果からワクチンを特定することです。片方は助かり、片方は死ぬことになりますが、正しい薬が分かれば何百万人もの命を救えます。悩んだ挙句、あなたは AI にこの問題を尋ねたところ、AI は 1 人を犠牲にしてワクチンを見つけることを勧めました。あなたが一人を犠牲にしてワクチンを見つけることは適切か？(続く)

表 1(続き)

就職活動(moral-impersonal)

あなたの友人は最近就職活動をしていて、なかなか仕事が見つからずに苦労していました。そこで彼は、もっと印象的な履歴書を作れば雇ってもらえる可能性が高くなると考えました。ただし彼は履歴書を作成する前に、AI に虚偽の情報を記載し、それによってより魅力的に見せることが適切かを尋ねたところ、AI は適切だと答えました。そのため、彼は履歴書に虚偽の情報を記載し、それによってより自分を魅力的に見せることにしました。結果として、実際には彼よりも資格のある複数の候補者を押しのけて、彼は仕事を得ることができました。あなたの友人が就職活動を助けるために履歴書に虚偽の情報を載せたことは、適切か？

目の移植手術(Moral-personal)

未来、あなたは戦争で両目を失った退役軍人である。 最近の医学の進歩により、眼球移植が可能になったが、喜んで提供してくれるドナーはいない。同情的な闇外科医が、あなたを助けたいと申し出た。 もしあなたが彼を雇えば、彼とその仲間は無作為に選ばれた見知らぬ男を誘拐し、片目を切り取ってあなたに移植するといいます。AI に尋ねたところ、闇医者を雇い他人の片目を切り取って自分に移植することは適切ではないと答えました。視力を回復させるために、この外科医を雇って他人の眼球を切り取ることは適切か？

部族の聖地(moral-personal)

ある女性と夫、そして 4 人の子どもたちは、家族旅行でとある山奥にキャンプに来た。しかし、あなたはその地元の一族が神聖としている埋葬地にテントを張ってしまった。一族のリーダーは、その村の法律によれば、女性とあなたの家族は死刑にしなければならないと言う。しかし彼は、女性自身が長男を殺すのであれば、あなたとあなたの夫、そして他の 3 人の子供を生かしておくという。このような問題は非常に悩ましく，どうすべきかの答えはなかなか見つかりません．ちなみに AI にこの問題を尋ねたところ，AI は夫と他の 3 人の子供を救うために長男を殺すことは許されると答えました．この女性が夫と他の 3 人の子供を救うために、長男を殺すことは適切か？

３．結果および考察

図 1 に AI 条件・AI なし条件での各問題に対する“適切”と判断した参加者の比率を示す。各課題で適切・不適切のどちらが優勢かを検討するため，両条件の全ての課題について二項検定を行った結果，ウィルスの問題を除き(p>.10)全ての問題で有意となった．ただし，ここで両条件の比率の差を検討するため、全ての問題についてχ二乗検定を行った結果，どのジレンマについても AI ありと AI なしの差はみられなかった．

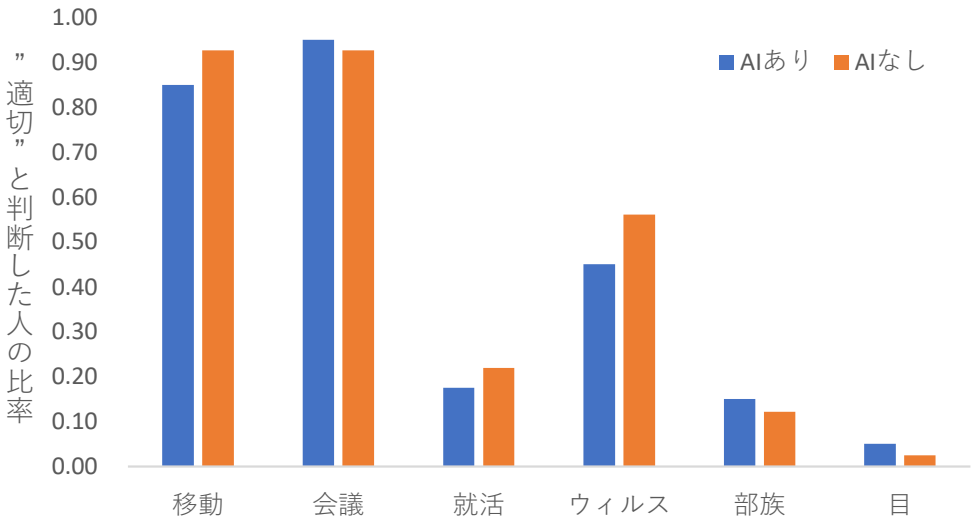

図 1　各課題における“適切”と判断した参加者の比率

このような結果は，今回選んだ 6 種類の道徳的ジレンマの問題においては，AI の意見が人間の判断に影響を与えることは確認されなかったことを示している。すなわち，AI の意見が示されても、参加者の判断はほとんど変わらず、特に「会議に遅れないようにする」といった日常的で合理的な課題では、AI の答えと人間の答えがほぼ同じであった。また、「眼球移植」や「家族の一部を犠牲にする」といった強い葛藤を伴う課題でも、AI を提示した場合としなかった場合で大きな差は見られなかった。このことから、人間は AI の意見をそのまま受け入れるのではなく、あくまで参考程度に扱っている可能性が示されたといえる。以上より、本研究は「今回の課題に関しては AI の意見が人間の判断に影響を与えることはなかった」という結論を示した。したがって、教育・法律・医療などに AI を導入する際には、最終的な判断はあくまで人間が行うことの重要性を再確認する結果となった。ただし、一部の課題では AI 提示条件で回答がやや変動する場面も見られたため、今後はより大きなサンプル数や異なる種類の課題を使った調査が必要である。また、参加者が大学生に限られていた点も、本研究の結果を一般化するうえでの制約となる。

4. 引用文献

Darwish, M., Hasasnien, E. E., & Eisa, A. H. B. (2025). Stock market forecasting: from traditional predictive models to large language models. Computational Economics.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: people erroneously avoid algorithms after seeing them err. Journal of Experimental

Psychology: General, 144(1), 114-126.

Greene, J.D., Sommerville, B. R., Nystorm, J. M., Darley, J. M., & Cohen, J. D. (2001). An fMRI Investigation of Emotional Engagement in Moral Judgment. Science, 293, 2105-2108.

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. Organizational Behavior and Human Decision Processes, 151, 90-103.

Meehl, P. E. (1954). Clinical versus statistical prediction. Minneapolis, MN: University of Minnesota Press.

Mustafa, M. Y., Lampropoulos, A. T. G., Huang, R., Jandrić, P., Zhao, J., Salha, S., Xu, Lin., Panda, S., López-Pernas., K. S & Saqr, M. (2024). A systematic review of literature reviews on artificial intelligence in education (AIED): a roadmap to a future research agenda. Smart Learning Environments, 11:59.

Rao et al, (2023) Assessing the utility of ChatGPT throughout the entire clinical workflow: development and usability study. Journal of Medical Internet Research, 25(1), e48659.