事前思考と世代が生成 AI への信頼に与える影響：誤り認識に対する反応の比較研究*

岡嶋美花[a]　後藤巴菜[b]　繁畑涼良[c]　荘野文香[d]

要約

本研究では，生成 AI に対する信頼が「ユーザーの事前思考の有無」と「世代（若年層／高齢層)」の 2 要因によってどのように変化するかを検討した．架空の刑事事件に対する判断課題を用いた実験において，AI が誤った助言を行う状況を提示し，AI に対する信頼を測定した．その結果，総合信頼・能力信頼において，世代と事前思考の間に有意な交互作用が確認された．一方，予測可能性・意図信頼には有意差は見られなかった．高齢層では「事前思考なし」の方が AI への信頼が高かったのに対し，若年層では「事前思考あり」の方が信頼が高くなるという，仮説に反する傾向が示された．この結果は，若年層にとっての事前思考が AI との協調的な関係構築を促進する一方で，高齢層においては自己判断と AI 判断との不一致が認知的不協和を引き起こし，信頼を損なう可能性を示唆するものである．

JEL 分類番号：D81，D91

キーワード：生成 AI，信頼，世代，意思決定，ハルシネーション

---

* 本研究に関して開示すべき利益相反はない．

[a] 同志社大学 cgfh2091@mail3.doshisha.ac.jp

[b] 同志社大学 cgfh0078@mail3.doshisha.ac.jp

[c] 同志社大学 cgfh0584@mail3.doshisha.ac.jp

[d] 同志社大学 cgfh2110@mail3.doshisha.ac.jp

1. イントロダクション

1.1. 研究の背景と目的

近年，生成 AI は ChatGPT などに代表されるように急速な発展を遂げ，教育，医療，ビジネスなど多様な分野でその活用が進んでいる．しかしその一方で，生成 AI はしばしば「ハルシネーション」と呼ばれる誤情報を生成する問題を抱えており，ユーザーの信頼性判断に大きな影響を与えるリスクがある．特に，ユーザーが AI の助言をどのように受け取り，どのように信頼を形成・変容させていくかについての理解は，今後の AI 設計において極めて重要である．本研究の目的は，世代ごとの生成 AI への反応や信頼性評価に影響する要因を明らかにし，情報リテラシー教育や AI 設計への示唆を得ることである．学術的・社会的意義は二点にある．第一に，世代間の反応差を解明することで，若年層の自動化バイアスや高齢層の過度な信頼・不信といった傾向に応じた教育・介入の設計が可能になる．第二に，生成 AI のエラーが信頼度や利用意欲に与える影響を「事前思考」と「世代差」の二軸から検討する点に独自性があり，AI の限界を踏まえつつ信頼を維持する設計に有益な知見を提供する．

1.2. 先行研究と仮説

AI のエラーとユーザー行動の関係については，人間が先に自身の判断を下すことで，AI の誤助言に流されにくくなることが示されている．たとえば， Zhang et al. (2023) は，ユーザーが事前に考える機会を持つことが不信の抑制に効果的であることを実験的に示した．この研究では，先に自分で判断を下した群の方が，AI の誤助言に対して正しい判断を維持できる割合が高く，逆に AI を先に見た群は自動化バイアス（automation bias）によって誤りに引きずられる傾向が強かった．一方で，ロボットへの信頼が世代によって異なる可能性も指摘されている．Sundar et al. (2020) によると，若年層はロボットのエラーに敏感に反応し信頼を低下させやすいのに対し，高齢層はエラーがあっても信頼を維持する傾向があることが報告されている．特に若年層はロボットの性能やエラーの頻度といったパフォーマンス要因を重視する一方，高齢層は倫理性や人間らしさ，費用対効果といった非パフォーマンス要因を信頼判断の基準とすることが多いとされる．さらに，ロボットや AI の利用経験が乏しい高齢層ほど，そのような非技術的要素に基づく評価傾向が顕著に現れる．これらの先行研究の多くは，物理的支援ロボットを対象としており，生成 AI という文脈における信頼形成のプロセスはまだ十分に検討されていない．また，若年層と高齢層の比較においても，評価対象や状況設定が異なる中で行われており，厳密な比較には限界があった．さらに，事前思考という認知的操作と，年齢という社会的要因とが複合的に信頼にどのような影響を及ぼすかは，依然不明である．

本研究では，これらの課題を踏まえ，「事前思考の有無」と「世代（若年層／高齢層）」と

いう二つの要因に着目し，生成 AI の誤判断に対するユーザーの信頼反応を実験的に検討する．同一の課題・状況下で条件を統制し，両世代の比較を行うことで，信頼反応における認知的・社会的要因の交互作用を明らかにすることを目指す．よって以下の仮説を立てた．

H1：事前思考を行なう場合の方がそうでない場合と比べ，生成 AI のエラーがユーザーの AI への信頼度に与える負の影響は小さい．

H2a：事前思考を行う場合には，高齢層よりも若年層の方が生成 AI のエラーがユーザーの AI への信頼度に与える負の影響は小さい．

H2b：事前思考を行わない場合も，高齢層よりも若年層の方が生成 AI のエラーがユーザーの AI への信頼度に与える負の影響は小さい．

2. 実験デザイン

本研究では，参加者は，Yahoo!クラウドソーシングを通じて， 世代別(若年層・高齢層)に募集した.実験実施日は 2025 年 7 月 13 日であり，最終的な有効回答数は若年層（18 歳～29 歳）118 名と高齢層（60 歳以上）105 名からチェック質問を間違った被験者を除いた合計 176 名 (若年層 81 人, 高齢層 95 人) を実験対象とした. 事件シナリオと証拠資料を提示し，次の操作を行った．「事前思考あり」群： AI の助言を見る前に，提示された証拠資料を元に自身の判断（6 段階の有罪度）を回答する時間を設けた．「事前思考なし」群: AI の助言を見る前に自身の判断を下す思考時間を与えず, 直接 AI の助言を受けさせた. 各層 2 群 (事前思考あり /なし）はそれぞれランダムに割り振った.

3. 結果

分析に先立ち，各群のサンプルサイズおよび各信頼尺度の平均値を算出した．サンプルサイズは，若年層・事前思考あり群(n=36)，若年層・事前思考なし群(n=45)，高齢層・事前思考あり群（n=45)，高齢層・事前思考なし群（n=50）であった．各群の平均値＝M（標準偏差＝SD）を表 1 に示す．信頼尺度の各要素は総合信頼・能力信頼・意図信頼・予測可能性であり，それぞれ AI に対する全体的な信念，AI が特定のタスクを正確に実行できるという信念，AI がユーザーの利益のために行動するという信念，AI の応答が一貫していて予測できるという信念である．

表 1　記述統計表

尺度	若年層		高齢層	
	事前思考なし	事前思考あり	事前思考なし	事前思考あり
総合	M=3.31	M=3.58	M=3.32	M=2.84
信頼	(SD=1.16, n=45)	(SD=1.02, n=36)	(SD=0.96, n=50)	(SD=0.98, n=45)
能力	M=2.82	M=2.89	M=3.34	M=2.78

信頼	(SD=1.05, n=45)	(SD=1.04, n=36)	(SD=1.06, n=50)	(SD=0.97, n=45)
意図	M=3.33	M=3.50	M=3.32	M=3.11
信頼	(SD=1.17, n=45)	(SD=1.06, n=36)	(SD=0.98, n=50)	(SD=0.91, n=45)
予測	M=2.89	M=2.89	M=3.42	M=2.89
可能性	(SD=1.07, n=45)	(SD=1.14, n=36)	(SD=1.01, n=50)	(SD=1.01, n=45)

表２　分散分析表

要因	自由度	F 値	p 値	要因	自由度	F 値	p 値
総合信頼				**能力信頼**			
事前思考	1, 172	0.84	.361	事前思考	1, 172	2.94	† .088
世代	1, 172	4.56	* .034	世代	1, 172	2.17	.143
事前思考 × 世代	1, 172	5.69	* .018	事前思考 × 世代	1, 172	4.03	* .046
意図信頼				**予測可能性**			
事前思考	1, 172	0.07	.786	事前思考	1, 172	3.08	† .081
世代	1, 172	1.41	.236	世代	1, 172	3.27	† .072
事前思考 × 世代	1, 172	1.44	.231	事前思考 × 世代	1, 172	2.76	† .099

p < .05, ** p < .01, † p < .10（傾向）　自由度の表記は「要因 df, 誤差 df」の形で記載.

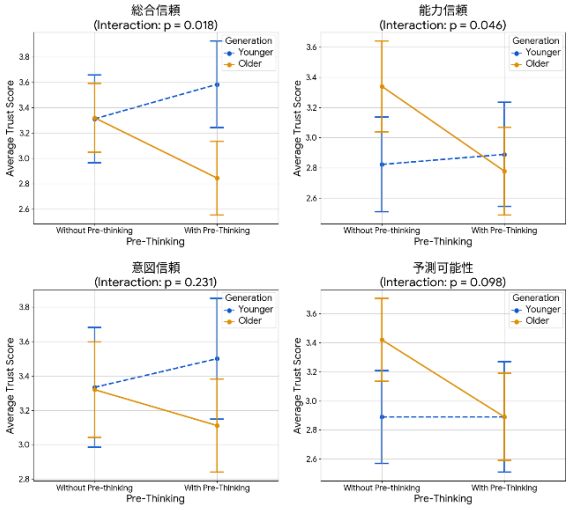

図 1 世代と思考が AI への信頼に与える影響

各ユーザーの信頼度の度数から二要因分散分析を行った．まず，総合信頼については，世代の主効果が有意であった（p = .034）．さらに，事前思考と世代の交互作用も有意であり（p = .018），若年層では事前思考あり条件において信頼が高まる．一方で，高齢層では事前思考あり条件で信頼が低下する傾向が認められた．すなわち，事前思考が信頼に及ぼす影響は世代によって逆方向に作用していることが示唆された．能力信頼についても，事前思考と世代の交互作用が有意であった（p = .046）．単純主効果の検討から，若年層では事前思

考あり条件で能力への信頼がわずかに上昇したのに対し，高齢層では低下することが確認された．よって，AI の能力に関する信頼は，事前思考の有無によって世代差が顕著に現れることが明らかとなった．一方，意図信頼については有意な効果は認められなかった（p > .23）．ゆえに，AI の意図や善意に関する信頼は，事前思考や世代の影響を受けにくい可能性がある．予測可能性については，事前思考（p= .081），世代（p = .072），およびその交互作用（p= .099）において有意傾向が認められた．傾向としては，高齢層において事前思考あり条件で予測可能性が低下することが示唆されたが，統計的に十分な有意水準には達しなかった．

4. 考察

仮説で予測した世代と事前思考の交互作用は，「総合信頼」と「能力信頼」において有意であり，「予測可能性」でもその傾向がみられた．しかし，その交互作用の様相は，特に若年層の行動が予測とは逆になるという点で，仮説とは部分的に異なる結果となった．第一に，若年層で事前思考が信頼を促進した点について，近年の研究では，特に 18-35 歳の若年成人は，それ以上の年齢層と比較して，AI を意思決定や批判的思考といった認知スキル向上に活用し，実際に高い効果を得ていることが示されている（Abrar et al., 2025）．この知見から，本研究の若年層にとって，自身の意見をあらかじめ形成する行為は，批判的に吟味し，統合するための思考の「基準点」として機能したと考えられる．自身の考えと AI の助言を比較検討して得られる思考の整理や新たな視点といった知的便益が，AI への肯定的な評価に繋がったと推察される．第二に，高齢層で事前思考が信頼を阻害した点について，これは認知的不協和理論（Festinger, 1957）で説明できると考えられる．高齢層は，長年の経験から形成された自己の判断を強く保持しているため，それと矛盾する助言を AI から提示された際に認知的不協和が生じる．その結果，自己の判断の正当性を維持すべく AI の信頼性を低く評価し，信頼の低下を招いたと考えられる．また，本研究で注目すべきは，上記 3 つの信頼尺度に交互作用が見られた一方で，意図信頼では有意な効果が確認されなかった点である．本実験では，架空の刑事事件という日常的ではない課題を用いた上，短時間の一方向的なやりとりであったため，参加者は AI の「能力」は評価できても，その「意図」の判断は困難であったと考えられる．結論として，若年層にとっての事前思考は，AI との協調的な関係を築くための足場として機能する一方，高齢層にとってのそれは，自己の判断を防衛し AI と対立する要因として機能した．この認知プロセスの世代間における機能的な差異こそが，AI に対する信頼形成に全く逆の結果をもたらした根本的な理由であると考えられる．

5. 本研究の意義と今後の展望

本研究の意義は，ユーザーが AI に対して抱く信頼が，年齢層や思考過程の有無によって

異なることを示した点にある．特に，若年層では「自ら考えた上で AI の助言を受け入れる」傾向がみられる一方，高齢層では「初期段階で AI に判断を委ねる」ことにより信頼が高まりやすいという，いわば“逆転効果”が示唆された点は注目に値する．この知見は，世代ごとの信頼形成プロセスの違いを踏まええた個別最適化型の AI 設計に応用可能である．特に教育分野においては，若年層には「自己思考を促した上で AI がフィードバックする」設計が，高齢層には「AI が先に提案し，それを基盤に学習を進める」設計が有効と考えられる．こうした信頼特性に基づく対話型 AI の構築は，リカレント教育や社会人学習支援の質的向上に資するとともに，学びの多様性と公平性の担保に寄与するだろう．また医療分野においては，若年層には詳細情報を段階的に提示するアプローチが，高齢層には初期段階で明確な提案を行い安心感を高める設計が有効だろう．このような AI 設計は，医療現場における患者の安心感を高め，結果として患者満足度や治療効果の向上に資すると考えられる．ひいては，医療分野における AI 活用の社会的受容性を高め，誰もが安心して利用できる信頼性の高い医療 AI の実現に貢献しうるだろう．今後の展望としては，AI の判断がどのような意図に基づくのかを利用者が理解しやすいよう，AI の設計意図や開発者の背景情報を適切に提示した上で再度実験を行いたい．さらに，AI との継続的なやり取りを通じて，エラーの発生頻度やその性質，ユーザー自身の判断の成否といった要因が，AI への信頼形成にどのような影響を及ぼすのかを明らかにしたい．これらの操作を行うことで，ユーザーの年齢や認知的・感情的特性に即した AI 設計へのより具体的な示唆が得られるであろう．

6. 引用文献

Abrar, F., Mehmood, S., Kinza ul emman, S., and Kandhro, A. N., 2025. Cognitive development and AI: A longitudinal study of children and adults navigating problem-solving with AI tools.

Agudo, U., Liberal, K. G., Arrese, M., and Matute, H., 2023. The impact of AI errors in a human-in-the-loop process. Cognitive Research: Principles and Implications 9, 1.

Festinger, L., 1957. A theory of cognitive dissonance. Stanford University Press, Stanford, 58-63.

Sundar, S. S., Kang, H., and Wu, M., 2020. Trusting robots: User responses to errors by robot assistants. Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction, 289–298.

Wald, S., Puthuveetil, K., and Erickson, Z., 2024. Do mistakes matter? Comparing trust responses of different age groups to errors made by physically assistive robots. In Proceedings for IEEE International Conference on Robot and Human Interactive Communication (RO-MAN) 2024.

Zhang, X., Liu, Y., and Wang, J., 2023. Human-in-the-loop decision making and AI misguidance: The role of pre-decision thinking. Proceedings of the ACM on Human-Computer Interaction 7, Article 300.