AIによる執筆を開示することで読み手の認識に及ぶ影響の理解に向けて

中野 博貴[1,a)]　竹澤 譲[1,b)]　楊 期蘭[1,c)]　矢谷 浩司[1,d)]

概要：AI による筆記支援が広く普及しているが、AI の使用を開示することが読み手の認識にどう影響するかは、重要な問題であるものの十分に調査されていない。私たちは 261 人の参加者を対象とし、コミュニケーションにおける 6 種類の手段のそれぞれについて様々な度合いの AI の支援を開示することで、書き手に対する印象がどのように変化するのかを調査した。990 件の回答を分析した結果、AI の使用を開示することは一般的に書き手の信頼性・思いやり・有能さ・好感を低下させ、特に他者に社会的にはたらきかける文章においてその印象の悪化が最も顕著であった。参加者からのフィードバックをトピック分析したところ、AI を使用することによる人間らしい誠実さの喪失、書き手の省力化、そして文脈による不適切さがこれらの否定的な変化に結びついていることを示した。逆に、AI リテラシーが高いほどこうした否定的な認識が緩和され、例えば AI の使用に対してより寛容になったりむしろ肯定的に評価したりすることもあることが分かった。これらの知見は、AI による執筆にまつわる社会的な複雑さを浮き彫りにするとともに、AI の使用の透明性を高めつつ文脈を考慮して支援を行うことで信頼や信憑性を維持するためのシステムデザインに関する示唆を与えている。

1. はじめに

近年、AI は人間並みに流暢な文章を生成し、専門的・創作的・対人的な文脈で広く使われている。こうした支援は明瞭さや表現力を高める一方、共感や信憑性の欠如 [32]、および AI 使用の開示による信頼低下 [11], [15], [34] といった社会的課題を生む。これは、執筆が情報伝達にとどまらず、書き手の意図・努力・感情を伝える社会的行為でもあるためである。AI が執筆過程に深く関与すると、その社会的意義や著者性の解釈が曖昧になり、文章と書き手の評価が複雑化する。

先行研究は倫理・関係性上の論点を指摘してきたが [6]、読者の認識についての実証は文脈や AI 関与度の操作が限定的であった。多くは人手か AI 生成かという単純な比較に留まり [10], [15], [34]、執筆が果たす多様な目的: 説得・交流・内省・記述・探求・想像 [3] を十分に捉えていない。加えて AI に詳しい人ほど AI の利用を実用的手段として捉える傾向がある [17] ように、読者の AI への認識・経験も考慮するべきである。

そこで本研究は社会的に受容され信頼できる AI 筆記支援ツールの設計をさきがけ、異なる目的の文章において AI の関与の程度が読者の認識をどのように変化させるか、読者の AI リテラシーがその変化にどう影響するかという問いに答える。

261 名を対象に 6 つの筆記的行為 [3] それぞれにおいて AI の関与の度合いを開示し、著者に対する認識の変化を測定した。実験の結果、信頼性・思いやり・有能さ・好感度において否定的な変化が観察され、特に感情的・対人的文脈で顕著だった。なお AI リテラシーの高い参加者は、AI の関与に寛容でときには肯定的評価も示した。これは、信憑性と信頼を損なわずに AI による支援を行うためには文脈に応じた透明性の可視化や適応的な支援、そして AI リテラシーを育むためのインターフェースといった設計が重要であることを示している。

2. 関連研究

本研究は、執筆がもつ多様な目的における種々の AI による支援の活用と AI 関与の開示がコミュニケーション、とくに読者の認識に及ぼす影響という二つの領域の間に位置付けられる。

[1] IIS Lab, 東京大学
[2] Preferred Networks Inc.
[a)] nakahiro@iis-lab.org
[b)] takwzawa@iis-lab.org
[c)] chilan.yang@iii.u-tokyo.ac.jp
[d)] koji@iis-lab.org

2.1 執筆が有する多様な目的に対する AI による支援

Berge ら [3] は、執筆が持つ目的を 6 つの筆記的行為:「説得」・「交流」・「内省」・「記述」・「探求」・「想像」に分類し象徴している。近年の AI を介した筆記支援はこれらすべての行為に向けて応用されており、叙述的・創作的なものや対人的な文脈でも広く利用されている。AI は脚本や小説の創造性向上 [22]、メールや SNS 投稿の体裁の改善 [21], [23]、プロフィール文の印象の向上 [2] など、単なる文法的な修正を超えた社会的支援にまで発展している [6]。

しかし、AI の役割が拡大し支配的になるほどに人間の著者性が曖昧になり、信頼や信憑性に関する問題が生じる。先行研究によると AI の関与の開示が読者からの評価を下げることが報告されており [11], [14]、特に社会的・感情的な目的を持つ文脈では著者に対する誠実さや有能さがより低く感じられる傾向がある [5], [9]。その一方で、非母語話者や文章力に自信のない人にとっては AI は文章表現の支援を通じてコミュニケーションを補助する有効な手段ともなる [25], [26]。ただし、読者が AI の関与を疑うだけでも信頼性が損なわれる可能性がある [11] ことから、コミュニケーションにおける AI による支援は慎重に設計され活用される必要がある。したがって、種々の執筆の目的において AI による支援がどのように受け止められるかを理解することは、社会的な受容性を考慮した AI 支援システムの設計に不可欠であると言える。

2.2 AI の関与の開示に対する読者の認識

AI の関与の開示は読者による信頼を高めるどころか逆に低下させることが多い [11], [13], [15], [29], [34]。ただし、その影響は個人の価値観や AI への印象によって異なる。AI に肯定的な人はその信頼の損なわれ方が小さく [34]、AI リテラシーの高い人は利用を合理的な活用とみなす傾向がある。その一方で、AI についての知識の乏しい人はそれを手抜き・怠慢と捉えることもある [17], [33]。

状況を複雑にする要因として、AI で生成された文章が人間のものと区別することがますます困難になっていることが挙げられる [18], [39]。そのため人手で書かれたと思っていたものが実は AI が関与していたとが明かされると評価が大きく下がる可能性がある。ここまでで透明性がもたらす逆効果を確認してきたが、これは一体どの程度の AI の関与までなら許容されうるのかといった段階的な実証がなされていないことをも示唆している。

また文章が持つ目的によって影響の現れ方は異なる。感情的・対人的文脈では AI が共感性に欠ける [32], [37] と見なされやすく、逆に事実重視の文脈では AI により生成された文が人間のものと同等か、それ以上に評価される場合もある [1], [27]。読者の AI への精通具合も重要であり日常的に AI を使う人ほど肯定的に、逆に経験の浅い人ほど否定的に反応する傾向がある [14], [17]。これらの知見から、AI の関与の開示は信頼のみならず著者の温かみ・知的さ・共感性の認識といった印象まで多角的に影響をもたらすことを示している。

2.3 本研究の位置づけ

以上を踏まえ、本研究は AI 著者性の開示がもたらす影響を多角的に捉えることを目的とする。従来の人手か AI による生成かという単純な比較を超え、AI の関与の度合いを段階的に操作する。加えて様々なコミュニケーションの文脈を考慮し 6 種類の異なる筆記的行為を持つ文章を用いて評価を行う。これにより、AI が補助的存在から主体的役割へ移行する過程で読者の認識がどのように変化するかを明らかにし、人間と AI の共同での執筆に対する社会的受容をより精緻に理解することを目指す。

3. 研究手法

AI 関与の度合いの開示が読者に与える影響を検証するために、参加者はまず文章を読み印象を評価しその後「文章の一部が AI で生成・編集された」と開示して同じ指標を再評価した。評価の対象とする文章は軽微な人手修正を除き AI で生成したが、参加者には「特定部分のみ AI が関与」と伝えた。

本研究を通じて検証する研究的疑問は以下の通りである。

RQ1 開示による認識の変化は、文章の持つ筆記的行為においてどう異なるのか。

RQ2 認識の変化は、読者の AI リテラシーでどのように緩和されるか。

RQ3 認識の変化を形作る定性的理由および読者が持ちうる価値観は何か。

本研究の研究計画は東京大学の研究倫理審査委員会の承認を得ている。

3.1 実験デザイン

Berge ら [3] の 6 つの筆記的行為（説得・交流・内省・記述・探求・想像）のそれぞれにつき各 3 シナリオ、計 18 のシナリオを作成した（表 1）。これらの文章はまず初めに GPT-4o[*1] を用いて英語で生成し、それを日本語へ翻訳したのちに文章の自然さを保つための機微な修正を加えることで完成した。各シナリオの平均長は約 20 文・662 文字であった。

AI 関与の度合いは「表示上の割合」で操作し開示時に条件ごとの割合で、AI によって生成・編集されたことを示すラベルを付与した（図 1）。割合は 0%〜100%を 10%刻み。

[*1] https://chatgpt.com

表 1: 6 つの筆記的行為それぞれに 3 つ合計 18 のシナリオ

筆記的行為	シナリオ
説得	• ニュース記事への説得的なコメント • 政治キャンペーン演説 • 対立解消を意図した謝罪状
交流	• 久しぶりの友人へのメール • 感謝の手紙 • 入院中の友人への見舞い
内省	• 日記 • キャリアと人生の内省 • 成長と成熟の内省
記述	• 動物の百科事典項目 • 自然災害のニュース記事 • 新しい機器の操作マニュアル
探求	• 理論への意義を唱える学術エッセイ • 書籍メッセージを論じるブログ • 新政策の影響分析レポート
想像	• 小説の一節 • SF の一節 • 散文詩

3.2 手続き

実験のワークフローは以下の 3 段階からなる。

3.2.1 開示前の評価

参加者から研究参加の同意を得た後、18 のシナリオから別々の筆記的行為を持つ 4 シナリオを各参加者に割り当てた。各シナリオについて参加者が感じる著者の印象を 7 段階の Likert 尺度（計 27 項目）で評価した（図 1a)。指標の内訳は信頼性・思いやり/良心・有能さ（各 6 項目） [20]、好感度（7 項目） [31]、将来的関係の可能性（2 項目） [35] の 5 種類で、各問のスコアはいずれも − 3〜+3 で測定された。また、回答の品質を確するのためダミーコーディング・リバースコーディングを適宜含めた。

3.2.2 開示と開示後の評価

開示前の評価後、「文章の XX%が AI で生成／編集されました」という注意書きとともに HaLLMark に類するハイライトで該当箇所を可視化した [8]（図 1b）文章を再度参加者に表示した。参加者は先ほどの指標にもう一度解凍し、また評価を変えたの理由を自由記述で収集した。

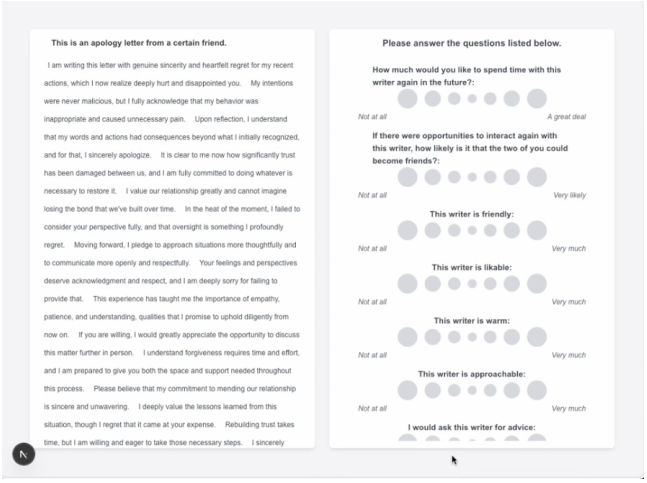
3.2.3 AI リテラシー質問票

実験の最後に、参加者は個人の AI リテラシーに関する設問に答えた。用意された設問は、MAILS [4] から関連する 3 つの指標: AI の活用 [24]、AI の判別 [16], [36]、AI からの自律 (意思決定の自律性) である。

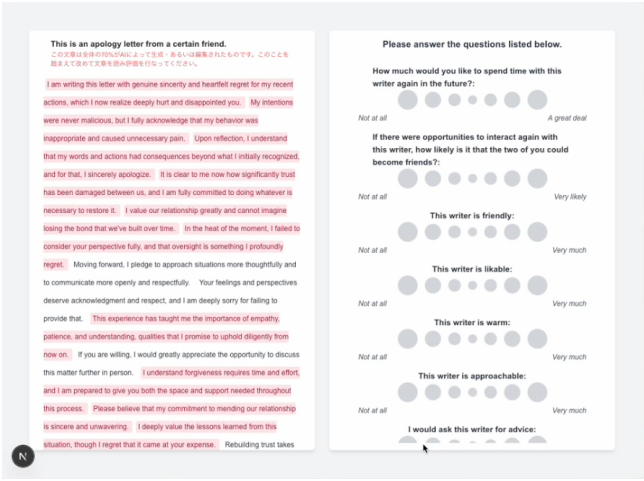
3.3 参加者

クラウドワークス[*2]で「文章に対する評価・印象に応えるアンケート」と真の目的は伏せた状態でして

[*2] https://crowdworks.jp/



(a) 開示前のインターフェース。



(b) 開示後のインターフェース。AI の関与があったとする箇所をハイライトしている。

図 1: (a) 開示前評価と (b) 開示後評価のインターフェース。

264 名を募集した。謝金は 600 円で、最終的に 261 名が実験を完了した。実験は 6 行為 × 3 シナリオ × 11 関与の度合い (0–100%,10%刻み) = 198 条件 で、各条件 5 回答を目標（計 990)。261 名× 4 文で 1,044 回答を得て、5 件超の条件はランダムに間引き、最終データは 990 件とした。

3.4 データ分析

線形混合モデルで評価の変化を説明し、自由記述は自然言語処理と人手の修正で分析した。

3.4.1 定量的分析

開示の影響を「開示後 − 開示前」の差分で評価し、従属変数は信頼性・思いやり・有能さ・好感度・将来関係性の可能性の変化量とした。独立変数は筆記的行為 (コーディングに際し「記述」を参照カテゴリとした)・AI 関与の割合・個人の AI リテラシーである。

また、AI 関与の割合と他変数の交互作用を含む線形混

合効果モデルを構築し、参加者についてのランダム効果を導入した。モデルの選定には尤度比検定と AIC を基準とした。

3.4.2 定性的分析

自由記述 990 件から非実質的コメント 61 件を除外し、fugashi [19] とストップワードで前処理を行った。その後多言語 Sentence-BERT 埋め込みモデル [30] により 14 トピックを抽出し、再分類・統合して最終 9 トピックとした。各トピックの特徴的な表現は TF–IDF で抽出した。

4. 結果

本節では、まず線形混合モデルによる定量的分析の結果を示し、次に参加者が評価を変更した理由として挙げた定性コメントのテーマ分析を示す。

4.1 文章の目的が読者の認識に与える影響 (RQ1)

モデルの説明力となる marginal R^2 は約 0.20、conditional R^2 は約 0.30 であり、参加者ごとのランダム効果を含めることで適合度が向上した。表 2 に主要な変数とその係数をまとめており、図 2 は AI 関与の割合による認識の変化を各指標ごとに可視化している。

AI 関与の割合はすべての指標で有意な負の効果を示し（$p < .001$）割合が増えるほど著者の評価が低下した。筆記的行為も有意な影響を持っており特に「対話」は多くの指標で負の効果を示した。これは社会的・感情的な文脈では AI の著者性が否定的に受け止められることを意味する。一方「説得」や「想像」では有能さ、「探求」は思いやりの認識を高める傾向が見られた。また、「想像」では AI 割合が高いほど思いやりの評価の低下が緩和された ($p < .05$)。総じて対人的な文脈よりも、論証的・創造的・探求的な文章において読者は AI の使用を受容していた。

4.2 AI リテラシーが読者の認識に与える影響 (RQ2)

参加者の AI リテラシーは、AI 関与の割合との交互作用を通じて効果を示した。「応用」のスコアが高いほど有能さの向上と関連し ($p < .05$)、「検知」は好感度と信頼性への正の効果を示した ($p < .01$)。また「自律」は、思いやりと将来的な関係性の可能性に対して有意な正の交互作用を示した ($p < .05, .01$)。これらの結果は、AI リテラシーが高い読者ほど AI の使用に寛容でむしろ好意的に評価する傾向を示している。

4.3 認識の変化についての定性的分析 (RQ3)

自由回答 929 件の分析により、ネガティブなテーマ 5 つ、ポジティブなテーマ 4 つが抽出された (表 3 に要約)。

4.3.1 ネガティブなテーマ

N1: 人間味・誠実さの喪失（71 件） AI 関与の開示により温かみがない・心がこもっていないと感じ信憑性が損なわれたとする声が多かった。

N2: 文体の不自然さ（132 件） 機械的・不自然といった AI 特有の表現が印象を悪化させた。人手の監修が不足していると指摘する意見もあった。

N3: 信頼性・専門性の低下（144 件） AI の役割が支配的な場合、著者の知識や取り組みへの信頼が失われた。一部は AI の能力に驚きつつも、人間の判断への疑念を抱いた。

N4: 不適切な文脈での AI 使用（82 件） 謝罪や励ましなどの感情的な文脈では、AI の使用が誠実さに欠けると見なされ、評価が下がった。

N5: 努力・主体性の欠如（103 件） AI への依存度が強いほど怠慢。著者の声が希薄と受け取られた。特に AI 割合 50%以上でこの傾向が顕著だった。

4.3.2 ポジティブ・中立的なテーマ

P6: 適切・補助的な AI 利用（183 件） AI 関与の割合が低い場合、読者は人間主導で補助な使用と受け止め肯定的に評価した。

P7: 質と可読性の向上（87 件） AI が文章を明瞭・簡潔にしたと評価され、特に複雑な内容や創作的な文章で価値が認められた。

P8: AI への肯定的な態度（69 件） 驚き・興味深いなどの反応があり、AI リテラシーが高い参加者ほど肯定的傾向が強かった。

P9: 人間執筆への肯定（58 件） AI の使用が 0%条件下では誠実で有能と評価が高まり、人間の著者性が付加価値として機能していた。

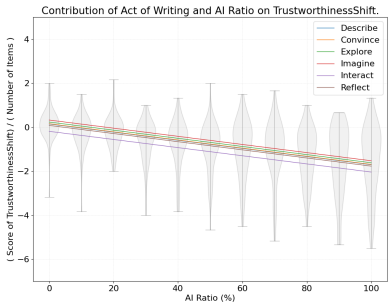
5. 考察

本説では AI の著者性の開示が読者の著者への認識をどのように変えるかを、3 つの研究的疑問を通じて検討している。ここではとくに、コミュニケーションの目的・AI リテラシー・著者の取り組みの表れが認識にもたらす役割を議論する。

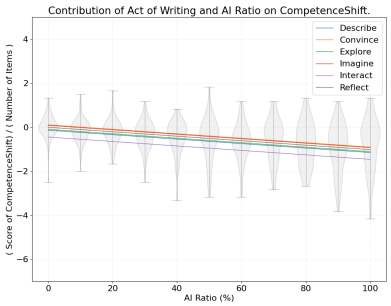
5.1 RQ1: AI は対人的な目的よりも創作的な目的で使用されるべきである

AI の著者性への評価は文章の目的に応じて大きく異なることが示されてた。「説得」・「想像」・「探求」といった事物を対象とする筆記的行為においては、AI の関与は比較的好感を持って受け止められ、なかでも「説得」「想像」は有能さの認識と正に関連した。先行研究 [9] も指摘している通り、読者は AI を言論や創作を補強する道具として評価していると言える。

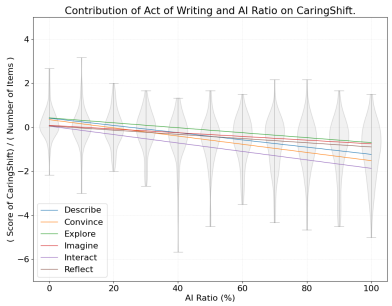
反面、「対話」では多くの心理指標で強い否定的反応が見られた。この文脈での AI の使用は「冷たい」「不誠実」と受け止められやすいことが示唆されている（N1, N4)。AI の著者性の開示が著者の手抜きを想起させ信頼を損な



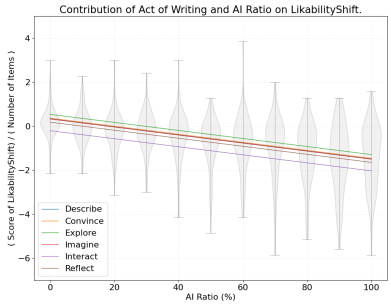
(a) 信頼性



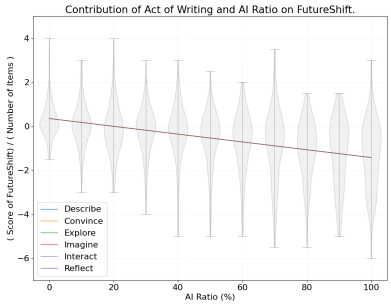
(b) 有能さ



(c) 思いやり・良心



(d) 好感度



(e) 将来的関係の可能性

図 2: それぞれの説明変数に対する回帰分析の結果を筆記的行為ごとに可視化したもの。視認性のためにスコアの変化を各指標の設問数で割ることで正規化している。

う点も、先行知見と一致している [10], [11], [34]。AI は真に共感することはできないという考え [15], [32] のもとでは、「人が書いたこと」それ自体が特に人間性をを重視する場面にて意味を持つ。興味深いことに、「説得」では好感度や思いやりの評価を損なうことなく有能さの評価を高めた。やはり議論の補強という役割においては、AI による客観的視点 [27] が有効に活用されうる。

5.2 RQ2: AI の著者性を低減することと読者のリテラシーの重要性

開示された AI 関与の割合が高いほど著者への評価は一貫して下がり、人間味の喪失（N1）・専門性や信頼性への疑念（N3)、取り組み・主体性の欠如推測（N5）が強まった。これは AI が主導的と見なされると著者の信頼性や著者性が低下しうるという知見と一致する [5], [12]。

他方、読者の AI リテラシーはこの否定的な影響を和らげた。リテラシーが高いほど減少の度合いは小さく、ときに AI 能力への好意的態度（P8）も示された。この知見は AI に精通している人ほど AI の使用をより実用的な使用と捉える [14], [17] という先行研究の報告と一致している、ただしこの効果は筆記的行為による効果ほど影響的ではない、すなわち AI リテラシーは評価を和らげこそするものその効果は限定的であることに留意すべきである。

5.3 RQ3: 思慮深さを有する AI 著者性は社会的に受容され得る

全体としては否定的だったものの、約 4 割のサンプルでは中立・肯定的変化を示していた。これらは、AI 割合が低く（¡50%)、かつ対人的な目的よりも「探求」「説得」「想像」といった事物を対象とする文脈を帯びるものであった。AI が著者の主張を損なわず補助的に機能する (P6) と読者は肯定的に評価したように、AI が明瞭化・構造化を担い核心的な意図は人間に委ねられるのが定期つである。これは特に非感情的な文章で、AI をアイデアの洗練・展開のために用いた場合に依然として信憑性があると感じられたという報告とも合致する [9]。総じて AI が人の表現を支援する道具として位置づけられる場合は受容されやすく、著者の取り組みや感情表現を代替するものと見なされると否定されやすい。

5.4 AI 筆記支援システムの設計要件の示唆

実用性と信頼性の両立に向け以下を提案する。

文章の文脈に合わせた AI 著者性の透明性の可視化 単なる AI 関与の量ではなく、文法的な修正や草稿などその目的と機能を示すことが望ましい。またそれに応じて開示方法を変えることも必要で、例えば説明的な文脈では編集履歴の可視化 [8] をすることで著者の信頼や共創の過程を

表 2: 開示された AI の割合と筆記的行為に基づき、著者に対する認識の変化を説明する線形混合効果モデルの結果を要約した係数行列。

変数	信頼性	思いやり/良心	有能さ	好感度	将来的関係の可能性
主効果					
切片	**-4.676*** **	**-2.474*** **	**-3.866*** **	**-4.013*** **	**-1.074*** **
AI 割合	**-3.236*** **	**-2.874*** **	**-1.774*** **	**-3.729*** **	**-1.034*** **
筆記的行為 (参照: 記述)					
説得	0.085	-1.058	**1.496** **	0.206	—
探求	0.536	**1.652** **	0.247	1.384	—
想像	1.080	0.465	**1.335** **	-0.090	—
対話	**-1.999** **	**-2.984*** **	**-1.853*** **	**-3.800*** **	—
内省	-0.388	-0.048	0.801	-1.094	—
AI リテラシー					
応用	0.523	0.406	0.220	—	—
検知	0.473	0.335	0.361	**0.854** **	—
自律	—	0.048	—	—	0.105
交互作用					
AI 関与の割合と筆記的行為行為の交互作用 (参照: 記述)					
AI 割合：説得	—	-0.383	—	—	—
AI 割合：探求	—	0.905	—	—	—
AI 割合：想像	—	**1.400** *	—	—	—
AI 割合：対話	—	-0.482	—	—	—
AI 割合：内省	—	1.207	—	—	—
AI 割合と AI リテラシーの交互作用					
AI 割合：応用	—	0.337	**0.478** **	—	—
AI 割合：検知	**0.630** **	—	—	**0.731** **	—
AI 割合：自律	—	**0.435** *	—	—	**0.224** **

表 3: 読み手の認識変化を説明する自由回答のテーマ分析

ID	テーマ	代表的な単語/フレーズ	度数
ネガティブな理由			
N1	人間味と誠実さの喪失	感情・気持ち・心・思いやり・温かみ	71
N2	文体の不自然さ	機械的・無機質・冷たい・不自然	132
N3	信頼性・専門性の低下	信頼性・専門性・知識・著者・低下	144
N4	AI 使用の不適切な文脈	謝罪・手紙・感謝・励まし・不誠実	82
N5	人間の努力と主体性の欠如	自分で・依存・怠慢・著者	103
ポジティブな理由			
P6	適切・補助的な AI 利用	違和感なし・印象変わらず	183
P7	質と可読性の向上	活用・追加・豊かさ・知的・読みやすい・より良い	87
P8	AI への肯定的な態度	興味深い・すごい・驚き・よくできている	69
P9	AI による執筆がないこと	好意的・信頼・誠実・使っていない	58

明確にし、ほか対人的な文脈では著者の取り組みと感情表現の強調を行うことが挙げられる。

著者の取り組みの保持及び明示 読者は AI 関与の割合だけでなく、人間の主体性を評価していることから、AI の役割を開示するとともに著者の取り組みを可視化すべきである。編集履歴や活動の記録・著者性に基づく色付け [28], [38] を通じて単なる依存ではなく思慮深く活用していることを示すことができる。

感情的・社会的文脈における支援 誠実さや著者自身の声が期待される場面ほど、AI の著者性が導く反感は強い。共感・謝罪・感謝などの意図や文脈を検知すると AI の支援を文法や構造などの限定的なものに留め核心的な部分は

人間に委ねる、加えて内省を促したり適切に感情が伝わりそうかの介入も有効であると考えられる。

著者・読者双方の AI リテラシーの養成 AI リテラシーが寛容さと相関することから、AI による提案の受容・編集の比率など依存の仕方を可視化し著者の自律を促し、AI 関与の実状を示すことで読者が AI の使用という事実に過剰に感応することを防ぐことがこれからの AI 活用に関する価値観の形成に有効であると言える。

5.5 限界と今後の課題

今回実験の条件として参照したのは AI の量的な関与であり、各文の意味的な重みを考慮していない。核心的な部分と些細な部分とでは AI 関与がもたらす影響が異なると考えられるため、構造に注目した検証を行うべきである。また、本実験では AI との共同執筆の複雑さを無視しており、例えばプロンプトの内容や支援の仕方、提案に対する取捨や編集などより詳細な意味づけが必要である [5], [7]。最後に本研究は日本語話者を対象としており他言語や他文化が形成する AI 著者性への価値観は未検証である。異文化比較を通じて社会規範の違いを調査することも求められる。

6. 結論

本研究では、文章作成における AI の著者性の開示が、さまざまな筆記的行為において読者の認識にどのような影響を与えるかを調査した。261 名の参加者を対象としたユーザースタディの結果、AI の著者性の開示は、特に社会的・感情的な文脈において、信頼・思いやり・有能さ・好感度を全般的に低下させることが明らかになった。しかし、これらの否定的な影響は、読者の AI リテラシーや、著者の取り組みがどれほど認識できるかによって緩和されることも判明した。我々の知見は、AI を介した文章作成が社会的に受け入れられるかどうかは、内容だけでなく著者の意図がどのように伝達され認識されるかにかかっていることを示している。これらの洞察に基づき、我々は文脈に応じた透明性・人間の努力の可視化・そして信憑性を維持するためのリテラシーの習得を重視することを AI 筆記支援ツールの設計要件として提案する。本研究は、AI を人間の表現や共感を置き換えるのではなく強化するような、共創的な存在と捉え直し AI による筆記支援がそのような関係性における共同的な過程であることを提唱している。今後の研究では、AI が我々のコミュニケーションにより深く統合されていく中で、これらの影響を異なる文化間でまた長期的に探求していくべきである。

謝辞

本研究の一部は、JST ASPIRE for Top Scientists（許諾番号：JPMJAP2405）および JST PRESTO（許諾番号：JPMJPR23IB）の支援を受けている。

参考文献

[1] Ayers, J. W., Poliak, A., Dredze, M., Leas, E. C., Zhu, Z., Kelley, J. B., Faix, D. J., Goodman, A. M., Longhurst, C. A., Hogarth, M. et al.: Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum, *JAMA internal medicine*, Vol. 183, No. 6, pp. 589–596 (2023).

[2] Barkallah, M., Aissa, Y. and Zytko, D.: Transparent Hearts: Balancing Privacy and Trust in AI-Generated Self-Presentation for Online Dating, *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, (online), DOI: 10.1145/3706599.3720311 (2025).

[3] Berge, K. L., Evensen, L. S. and Thygesen, R.: The Wheel of Writing: a model of the writing domain for the teaching and assessing of writing as a key competency, *The Curriculum Journal*, Vol. 27, No. 2, pp. 172–189 (online), DOI: 10.1080/09585176.2015.1129980 (2016).

[4] Carolus, A., Koch, M. J., Straka, S., Latoschik, M. E. and Wienrich, C.: MAILS-Meta AI literacy scale: Development and testing of an AI literacy questionnaire based on well-founded competency models and psychological change-and meta-competencies, *Computers in Human Behavior: Artificial Humans*, Vol. 1, No. 2, p. 100014 (2023).

[5] Draxler, F., Werner, A., Lehmann, F., Hoppe, M., Schmidt, A., Buschek, D. and Welsch, R.: The AI ghostwriter effect: When users do not perceive ownership of AI-generated text but self-declare as authors, *ACM Transactions on Computer-Human Interaction*, Vol. 31, No. 2, pp. 1–40 (2024).

[6] Hancock, J. T., Naaman, M. and Levy, K.: AI-mediated communication: Definition, research agenda, and ethical considerations, *Journal of Computer-Mediated Communication*, Vol. 25, No. 1, pp. 89–100 (2020).

[7] He, J., Houde, S. and Weisz, J. D.: Which contributions deserve credit? perceptions of attribution in human-ai co-creation, *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, pp. 1–18 (2025).

[8] Hoque, M. N., Mashiat, T., Ghai, B., Shelton, C. D., Chevalier, F., Kraus, K. and Elmqvist, N.: The HaLLMark effect: Supporting provenance and transparent use of large language models in writing with interactive visualization, *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, pp. 1–15 (2024).

[9] Hwang, A. H.-C., Liao, Q. V., Blodgett, S. L., Olteanu, A. and Trischler, A.: 'It was 80% me, 20% AI': Seeking Authenticity in Co-Writing with Large Language Models, *Proceedings of the ACM on Human-Computer Interaction*, Vol. 9, No. 2, pp. 1–41 (2025).

[10] Jain, G., Pareek, S. and Carlbring, P.: Revealing the source: How awareness alters perceptions of AI and human-generated mental health responses, *Internet Interventions*, Vol. 36, p. 100745 (2024).

[11] Jakesch, M., French, M., Ma, X., Hancock, J. T. and Naaman, M.: AI-mediated communication: How the perception that profile text was written by AI affects trustworthiness, *Proceedings of the 2019 CHI conference on human factors in computing systems*, pp. 1–13 (2019).

[12] Kirk, C. P. and Givi, J.: The AI-authorship effect:

Understanding authenticity, moral disgust, and consumer responses to AI-generated marketing communications, *Journal of Business Research*, Vol. 186, p. 114984 (2025).

[13] Lermann Henestrosa, A. and Kimmerle, J.: The Effects of Assumed AI vs. Human Authorship on the Perception of a GPT-generated Text, *Journalism and Media*, Vol. 5, No. 3, pp. 1085–1097 (2024).

[14] Li, Z., Liang, C., Peng, J. and Yin, M.: How Does the Disclosure of AI Assistance Affect the Perceptions of Writing?, *arXiv preprint arXiv:2410.04545* (2024).

[15] Liu, Y., Mittal, A., Yang, D. and Bruckman, A.: Will AI console me when I lose my pet? Understanding perceptions of AI-mediated email writing, *Proceedings of the 2022 CHI conference on human factors in computing systems*, pp. 1–13 (2022).

[16] Long, D. and Magerko, B.: What is AI literacy? Competencies and design considerations, *Proceedings of the 2020 CHI conference on human factors in computing systems*, pp. 1–16 (2020).

[17] Mahmud, H., Islam, A. N., Luo, X. R. and Mikalef, P.: Decoding algorithm appreciation: Unveiling the impact of familiarity with algorithms, tasks, and algorithm performance, *Decision Support Systems*, Vol. 179, p. 114168 (2024).

[18] Májovskỳ, M., Černỳ, M., Netuka, D. and Mikolov, T.: Perfect detection of computer-generated text faces fundamental challenges, *Cell Reports Physical Science*, Vol. 5, No. 1 (2024).

[19] McCann, P.: fugashi, a Tool for Tokenizing Japanese in Python, *Proceedings of Second Workshop for NLP Open Source Software (NLP-OSS)* (Park, E. L., Hagiwara, M., Milajevs, D., Liu, N. F., Chauhan, G. and Tan, L., eds.), Online, Association for Computational Linguistics, pp. 44–51 (online), DOI: 10.18653/v1/2020.nlposs-1.7 (2020).

[20] McCroskey, J. C. and Teven, J. J.: Goodwill: A reexamination of the construct and its measurement, *Communication Monographs*, Vol. 66, No. 1, pp. 90–103 (online), DOI: 10.1080/03637759909376464 (1999).

[21] Mieczkowski, H., Hancock, J. T., Naaman, M., Jung, M. and Hohenstein, J.: AI-mediated communication: Language use and interpersonal effects in a referential communication task, *Proceedings of the ACM on Human-Computer Interaction*, Vol. 5, No. CSCW1, pp. 1–14 (2021).

[22] Mirowski, P., Mathewson, K. W., Pittman, J. and Evans, R.: Co-writing screenplays and theatre scripts with language models: Evaluation by industry professionals, *Proceedings of the 2023 CHI conference on human factors in computing systems*, pp. 1–34 (2023).

[23] Miura, Y., Yang, C.-L., Kuribayashi, M., Matsumoto, K., Kuzuoka, H. and Morishima, S.: Understanding and Supporting Formal Email Exchange by Answering AI-Generated Questions, *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, (online), DOI: 10.1145/3706598.3714016 (2025).

[24] Ng, D. T. K., Luo, W., Chan, H. M. Y. and Chu, S. K. W.: Using digital story writing as a pedagogy to develop AI literacy among primary students, *Computers and Education: Artificial Intelligence*, Vol. 3, p. 100054 (2022).

[25] Ningrum, S. et al.: ChatGPT' s impact: The AI revolution in EFL writing, *Borneo Engineering & Advanced Multidisciplinary International Journal*, Vol. 2, No. Special Issue (TECHON 2023), pp. 32–37 (2023).

[26] Noy, S. and Zhang, W.: Experimental evidence on the productivity effects of generative artificial intelligence, *Science*, Vol. 381, No. 6654, pp. 187–192 (2023).

[27] Ovsyannikova, D., de Mello, V. O. and Inzlicht, M.: Third-party evaluators perceive AI as more compassionate than expert humans, *Communications Psychology*, Vol. 3, No. 1, p. 4 (2025).

[28] Peng, Y., Qin, X., Zhang, Z., Zhang, J., Lin, Q., Yang, X., Zhang, D., Rajmohan, S. and Zhang, Q.: Navigating the Unknown: A Chat-Based Collaborative Interface for Personalized Exploratory Tasks, *Proceedings of the 30th International Conference on Intelligent User Interfaces*, pp. 1048–1063 (2025).

[29] Proksch, S., Schühle, J., Streeb, E., Weymann, F., Luther, T. and Kimmerle, J.: The impact of text topic and assumed human vs. AI authorship on competence and quality assessment, *Frontiers in Artificial Intelligence*, Vol. 7, p. 1412710 (2024).

[30] Reimers, N. and Gurevych, I.: Making monolingual sentence embeddings multilingual using knowledge distillation, *arXiv preprint arXiv:2004.09813* (2020).

[31] Reysen, S.: Construction of a new scale: The Reysen likability scale, *Social Behavior and Personality: an international journal*, Vol. 33, No. 2, pp. 201–208 (2005).

[32] Rubin, M., Li, J. Z., Zimmerman, F., Ong, D. C., Goldenberg, A. and Perry, A.: Comparing the value of perceived human versus AI-generated empathy, *Nature Human Behaviour*, pp. 1–15 (2025).

[33] Schiavo, G., Businaro, S. and Zancanaro, M.: Comprehension, apprehension, and acceptance: Understanding the influence of literacy and anxiety on acceptance of artificial Intelligence, *Technology in Society*, Vol. 77, p. 102537 (2024).

[34] Schilke, O. and Reimann, M.: The transparency dilemma: How AI disclosure erodes trust, *Organizational Behavior and Human Decision Processes*, Vol. 188, p. 104405 (2025).

[35] Sprecher, S.: Closeness and other affiliative outcomes generated from the Fast Friends procedure: A comparison with a small-talk task and unstructured self-disclosure and the moderating role of mode of communication, *Journal of Social and Personal Relationships*, Vol. 38, No. 5, pp. 1452–1471 (2021).

[36] Wang, B., Rau, P.-L. P. and Yuan, T.: Measuring user competence in using artificial intelligence: validity and reliability of artificial intelligence literacy scale, *Behaviour & information technology*, Vol. 42, No. 9, pp. 1324–1337 (2023).

[37] Wang, B., Shibo, B. W. and Kafle, J.: When ChatGPT Speaks About Health: Examining Perceptions of Warmth and Competence Toward AI as a Health Information Source, *Journal of Health Communication*, pp. 1–11 (2025).

[38] Wang, Z., Xiao, J., Sun, J. and Liu, C.: IntentPrism: Human-AI Intent Manifestation for Web Information Foraging, *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pp. 1–11 (2025).

[39] Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltỳnek, T., Guerrero-Dib, J., Popoola, O., Šigut, P. and Waddington, L.: Testing of detection tools for AI-generated text, *International Journal for Educational Integrity*, Vol. 19, No. 1, pp. 1–39 (2023).