

[研究助成 (B)]

Investigating the Effect of AI on Meaningful
Interpersonal Communication

2241902

研究代表者
(助成金受領者)

東京大学 情報学環

助教 楊 期 蘭

[研究の目的]

感情的に負荷の高いデジタルコミュニケーションの課題

職場のコミュニケーションでは、ネガティブな感情を引き起こす困難な会話が避けられない場面があります。テキストベースの非同期コミュニケーションでは非言語的な手がかりが欠如しているため、感情的に負荷の高いメッセージ（厳しい批判や困難な職場の指示など）は、意図された以上に攻撃的・不公平・士気低下をもたらすものとして受け取られることがあります。研究によれば、ネガティブなフィードバックは受信者のモチベーション、自己効力感、そして生産的に関与しようとする意欲を低下させる可能性が示されています。

査読においては、建設的でない批評が著者に研究を諦めさせたり、研究コミュニティから離れさせたりする原因となることがあります。職場では、攻撃的または素っ気ないメールがコンフリクトを激化させ、感情的なエネルギーを消耗させ、効果的なコミュニケーションを妨げます。

介入戦略としての認知的再評価

認知的再評価とは、状況を別の視点から解釈し直すことで感情的な影響を変える心理のプロセスであり、感情調整の有効な方略として確立されています。しかし、特にストレス下では認知的負荷が高く、通常は訓練や専門家のサポー

トが必要です。大規模言語モデル（LLM）は、困難なコミュニケーションに対して、ポジティブに再フレーミングされた代替的視点を自動的に生成することで、認知的再評価をより利用しやすくする有望な可能性を秘めています。

以前の研究ではメンタルヘルスの文脈でこの可能性が探られていましたが、本研究の2編は、AI駆動の再評価が自然なコミュニケーション場面（具体的には査読と職場コンフリクト）において果たす役割を厳密に検討した最初の試みの一つです。

研究目的

2つの研究は、以下の共通した目的によって結びついています。

- ・ 困難なメッセージに対する AI 支援のポジティブなリフレーミングが、受信者のネガティブな感情を低減し、メッセージおよび送信者に対する認識を改善するかどうかを検討すること。
- ・ AI リフレーミングの効果が、行動的アウトカム（査読における改訂の努力や、職場のコミュニケーションにおけるコンフリクト管理スタイルなど）にまで及ぶかどうかを調査すること。
- ・ AI による再評価の効果を調整する要因（例：総合評価スコア、AI の物語的視点）を理解すること。
- ・ 意図を歪めたり、主体性を損なったり、人間の責任を代替したりすることなく、人間の関

係的な効果を支援する責任ある AI ツールの設計に貢献すること。

[研究の内容, 成果]

研究 1－査読における AI 再フレーミング済みポジティブ要約 (CHI '25)

Prolific を通じて募集した 137 名の参加者を対象に、2 段階のオンライン統制実験を実施しました。第 1 段階では、参加者はソーシャルメディアマーケティング職への応募用カバーレター (250 語) を作成しました。このタスクは個人的な関与を引き出しつつも統制可能になるよう設計されました。3～7 日後の第 2 段階では、参加者は GPT-4 によって意図的に厳しいトーンで生成された AI フィードバックと、総合評価スコアを受け取りました。

本研究は 2×2 の被験者間デザインを採用し、(1) 批評に付加される AI 再フレーミング済みポジティブ要約の有無 (あり vs. なし) と、(2) 総合評価レベル (高: 5 点満点中 4 点 vs. 低: 5 点満点中 2 点) を操作しました。

ポジティブ要約は GPT-4 と認知的再フレーミングプロンプトを用いて生成され、元のフィードバックを置き換えたり和らげたりするのではなく、批評の中の建設的な機会を強調するよう設計されました。

アウトカム指標には、フィードバックの感情的価値の知覚、文章の自己効力感の変化、自律

性・有能感の知覚、フィードバックの公正さ・有用性の知覚、査読者の公正さ・専門性の知覚、改訂の量 (TF-IDF コサイン類似度で測定)、改訂の質 (ループブリックを用いた 2 名の独立した評定者によるスコア)、および自己評価による改訂努力が含まれました。自由記述式の調査回答は、動機づけの帰属理論を用いて主題分析し、改訂行動を促す個人内・対人的要因を特定しました。(表 1: 左)

研究 2－困難な職場メールにおける社会的緩衝としての AI (CHI '26)

CloudResearch を通じて募集した 132 名の参加者を対象に、オンライン統制実験を実施しました。参加者は 4 種類の困難な職場メールのうちの 1 つを受け取り、2×2 の要因計画によって設定された 5 つの条件のいずれかに割り当てられました。条件は、AI によるポジティブなリフレーミングとニュートラルなリフレーミング、および 1 人称と 3 人称の物語的視点の組み合わせです。AI リフレーミングなしの統制条件 (n = 27) も設けられました。すべてのリフレーミングされたメッセージは、実験前に ChatGPT-3.5 を用いて生成され、参加者間で一定に保たれました。

アウトカム指標には、感情反応 (Russell 感情円環モデルに基づく、価値と覚醒の 4 象限)、コンフリクト管理行動 (対人コンフリクト相互作用コーディングシステムを用いて符号化)、メール返信と解釈の言語的特徴 (LIWC-22 を用いて分析、特に権力関連語と認知プロセス言語に着目)、AI リフレーミングメッセージの知覚された有用性、および AI への信頼が含まれました。(表 1: 右)

結果

AI 支援のポジティブリフレーミングがもたらす感情的・知覚的恩恵

両研究を通じて、AI 支援のポジティブリフレーミングは一貫して受信者の感情的体験と対

表 1 2つの研究のデザインの概要

項目	研究 1 (CHI '25)	研究 2 (CHI '26)
状況	カバーレターの学術的な査読	困難な職場のメール会話
参加者数 (N)	137 名	132 名
デザイン	2×2 被験者間デザイン	2×2+ 統制群の被験者間デザイン
主な操作変数	AI ポジティブ要約 (あり/なし) × 評価スコア (高/低)	リフレーミングの種類 (ポジティブ/ニュートラル) × 物語的視点 (1 人称/3 人称)
アウトカムの焦点	感情的受容, 改訂行動, 動機づけ	感情, コンフリクト行動, 言語使用, AI への信頼

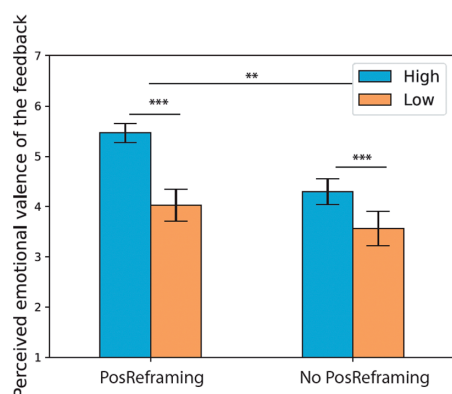


図1 各条件における感情的価値

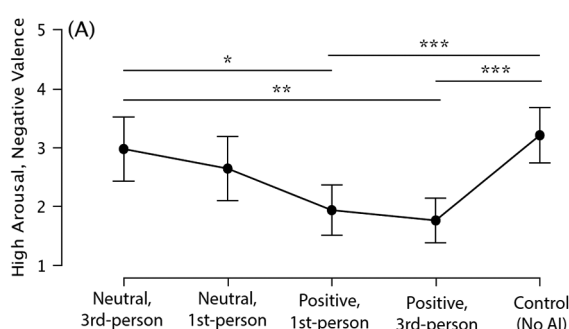


図2 各条件におけるネガティブな感情的価値の水準

人的認知を改善しました。

研究1では、厳しいフィードバックとともにAI再フレーミング済みポジティブ要約を受け取った参加者が、受け取らなかった参加者と比較して、有意にポジティブな感情的価値 ($p < .01$) (図1)、改訂における自律性と有能感の高い知覚 ($p < .01$)、フィードバックの公正さへの高い評価 ($p < .01$)、および査読者の公正さへの高い評価 ($p < .05$) を示しました。これらの効果は、総合評価スコアが高いか低いかにかかわらず見られました。

研究2では、ポジティブにリフレーミングされたメールを受け取った参加者が、ニュートラルリフレーミング条件や統制条件と比較して、有意に低い高覚醒ネガティブ感情 (例：ストレス、緊張; $p < .001$) を経験しました (図2)。また、ニュートラルリフレーミングと比較して、有意に低い低覚醒ネガティブ感情 ($p < .05$) も示されました。

全体として、ポジティブリフレーミングはネ

ガティブな感情を低減しますが、高覚醒のポジティブ感情を有意に高めるわけではありません。これはAIによる再評価が、ポジティブな感情を生成するというよりも、主に感情的緩衝として機能している——つまりネガティブな影響を和らげる——ことを示唆しています。

行動的アウトカム：従来の状態の持続性

感情的・知覚的な改善が明確に見られた一方で、両研究において行動的アウトカムはAIリフレーミングに対して反応しにくいことがわかりました。

研究1では、AIリフレーミング済みポジティブ要約は、参加者の文章改訂の量や質を有意に増加させませんでした。予想外なことに、改訂努力を高めたのはポジティブリフレーミングではなく、低い総合評価スコアでした ($p < .05$)。これは、知覚されたパフォーマンスギャップが自己改善の動機を高めるという研究と一致しています。また、ポジティブリフレーミングを受け取った参加者は、自分の改訂の質をより高く評価しました ($p < .05$)。

同様に、研究2においても、ポジティブなりフレーミングもニュートラルなりフレーミングも、参加者のコンフリクト管理行動やメール返信における送信者の承認に有意な変化をもたらしませんでした。返信の感情的トーンや語数も条件間で影響を受けませんでした。ただし、注目すべき言語的な知見が見られました。ポジティブリフレーミングを受けた参加者は、1人称のニュートラルリフレーミングを受けた参加者と比較して、困難なメールの解釈において権力関連語 (例：「命令」「統制」「権威」) を有意に少なく使用しました ($p < .01$, Cohen's $d = 0.91$) (図3)。これは、明示的な行動はすぐには変化しなくても、ポジティブリフレーミングが困難な対人関係における権力動態の認知的な枠組み方を、微妙に変える可能性を示しています。

AIへの信頼と知覚された有用性

研究2における注目すべき知見として、ポジ

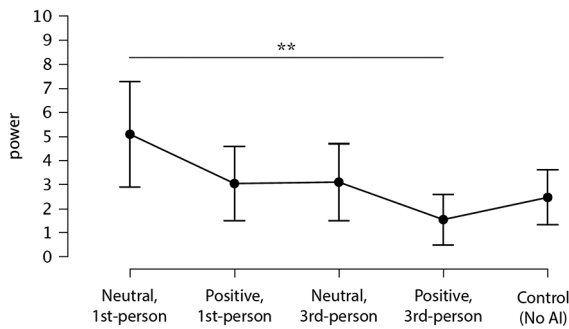


図3 各条件における権力関連語の使用頻度

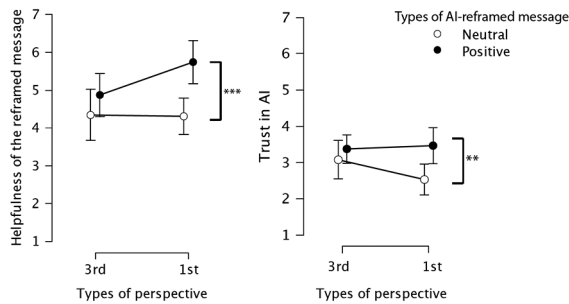


図4 各条件における知覚された有用性と信頼の水準

タイプにリフレーミングされたメッセージは、物語的視点にかかわらず、ニュートラルリフレーミングと比較して、知覚された有用性 ($p < .001$, Cohen's $d = 0.69$) と AI への信頼 ($p < .01$, Cohen's $d = 0.54$) の両方を有意に高めました (図4)。このことは、ユーザーが AI を評価する際、透明性や技術的な能力だけでなく、出力の質と温かみに基づいていることを示唆しています。この知見は、信頼の高まりが意図せず人間の送信者から AI へと責任を移行させる可能性を示唆しており、これは慎重な設計上の注意を要する課題です。

統合：共通テーマと設計への示唆

2つの研究を総合すると、一つの共通パターンが浮かび上がります。AI によるポジティブリフレーミングは、感情的・知覚的なレベルでは信頼性高く効果を発揮しますが、外顕的な行動を決定論的に変えるわけではない、というものです。

感情と行動の間にあるこのギャップは、理論的に重要な知見です。AI による再評価を通じた感情調整は、行動変容の必要条件ではあっても十分条件ではない可能性が示唆されます——明示的なコンフリクト解決戦略や具体的な改訂ガイダンスなど、追加的な支援が必要かもしれません。

設計者や研究者にとって、これらの知見は以下のことを示しています。人間のコミュニケーションを改善する AI ツールは、(1) ニュートラルなトーン調整よりもポジティブリフレーミングを優先すること、(2) AI が生成したコンテンツを元のメッセージの代替ではなく補助的なものとして提示すること、(3) AI の役割を全ての関係者に対して透明にすること、(4) 感情的サポートを建設的な行動のガイダンスで補完すること、(5) AI によるメッセージの和らげが有害なコミュニケーションを正常化したり送信者の意図を誤って伝えたりするリスクに注意を払い続けること、が求められます。

AI が人間の相互作用を媒介する場面がますます増えていく中で、これらの知見は意義ある対人コミュニケーションを可能にするシステムを設計するための実証的な基盤を提供します。

[成果の発表、論文など]

1. Chi-Lan Yang, Alarith Uhde, Naomi Yamashita, and Hideaki Kuzuoka. Understanding and Supporting Peer Review Using AI-reframed Positive Summary. In Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25), Yokohama, Japan, Article 171, 16p., doi:10.1145/3706598.3713219, 2025.4
2. Chi-Lan Yang, Jing Li, Xuhui Chang, Jingshu Li, Koji Yatani, Yi-Chieh Lee. Can AI be a Social Buffer? Investigating the Effect of AI-assisted Cognitive Reappraisal and Narrative Perspectives on Managing Difficult Workplace Conversations over Email. Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), Barcelona, Spain, Article 1514, 19p., doi.org/10.1145/3772318.3790331, 2026.4