

AIによる質問を利用したメール返信支援システムの評価

三浦 悠輔* 楊 期蘭† 栗林 雅希* 松本 啓吾† 葛岡 英明† 森島 繁生‡

概要. 長くて丁寧な職場のメールに返信することは、時間がかかり、認知的にも負担が大きい。大規模言語モデルを用いた返信作成支援システムは、メール返信プロセスを簡素化することを目的として提案されているが、ユーザは依然として期待する出力を得るために詳細なプロンプトを作成する必要がある。そこで本研究では、受信メールの内容を基に AI によって生成されたシンプルで短い質問への回答を通じて、ユーザが返信を作成できる、質問-回答方式のアプローチを提案し、評価した。本研究では、12 名と 9 名を対象に、実験室環境と実環境でユーザ実験を実施した。その結果、提案手法は従来のプロンプトベースのアプローチと比較して、ユーザが適切なプロンプトを作成する手間を省き、メールの品質を維持しながら、返信の効率を向上させ、また作業の負担を軽減することが明らかとなった。

1 はじめに

メールは、情報共有 [11, 13] やタスク管理 [3] に広く使われているツールだが、メールの確認や返信は時間がかかり、認知的な負担が大きい [9]。特に礼儀を重んじる文化では、相手の意図を汲み取り、丁寧なメッセージを作成する必要があるため、負担が増える。こうした負担を軽減するために、返信方法や内容を支援する様々なツールが提案されてきた。特に近年では、生成 AI の発展により、AI を活用したコミュニケーションツールが増えており、メール返信の下書きを生成するツールも多数登場している [12, 8, 10, 2, 1, 7, 5, 4]。しかし、これらのツールを効果的に使うためには、ユーザは適切なプロンプトを作成する必要がある。期待する結果が得られない場合にはプロンプトを繰り返し修正する必要がある [6]。この問題に対処するため、AI による質問を用いた質問-回答方式 (QA ベース) のアプローチ、ResQ を提案する。このアプローチでは、プロンプトの作成作業が、AI が生成する質問に対する回答作業に置き換えられることで、返信作業の負担を軽減することが期待される。

本研究の目的は、2 種類のユーザ実験を通じて、ResQ がユーザの作業や心理に及ぼす影響を定性的かつ定量的に評価することである。12 名の参加者を対象にした実験室環境での実験と 9 名の実環境での実験の結果、従来のプロンプトベースの方法と比較して、作業効率が向上し、作業負担が軽減することが確認された。またユーザの行為主体感は低下することも確認されたが、実環境での実験のインタ

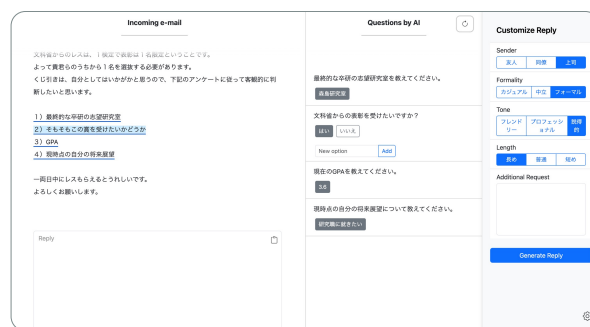


図 1. ResQ のインターフェース。

ビュー結果から、一部の参加者はコミュニケーションの質や量が向上したと認識し、心理的距離が縮まる可能性が示唆された。

2 提案アプローチ

図 1 は、実際の ResQ のインターフェースを示している。ResQ はまず、ユーザが返信作業を開始すると、大規模言語モデル (本研究では GPT-4o) を使用し、メールの本文や送信者情報、過去のやり取り、ユーザの情報を基に、多肢選択式の質問を生成する。ユーザはこれらの質問に、ResQ が生成した選択肢を選択するか、自身で回答を入力することで、ResQ に返信方針を伝える。その後、ユーザが「返信を生成」ボタンを押すと、ResQ は LLM を使ってユーザの回答と要望に基づいた返信の下書きを生成し、ユーザ提示する。最後に、ユーザは必要に応じて下書きを確認・修正し、返信を送信する。

このような QA ベースのアプローチは、ユーザの作業効率の向上、認知負荷の低減、タスクの満足度の向上、作業開始のハードルの低減、メールの品質の向上、コミュニケーション相手に与える印象の向上を目的としている。一方で、ユーザの行為主体感

Copyright is held by the author(s). This paper is non-refereed and non-archival. Hence it may later appear in any journals, conferences, symposia, etc.

* 早稲田大学

† 東京大学

‡ 早稲田大学理工学術院総合研究所

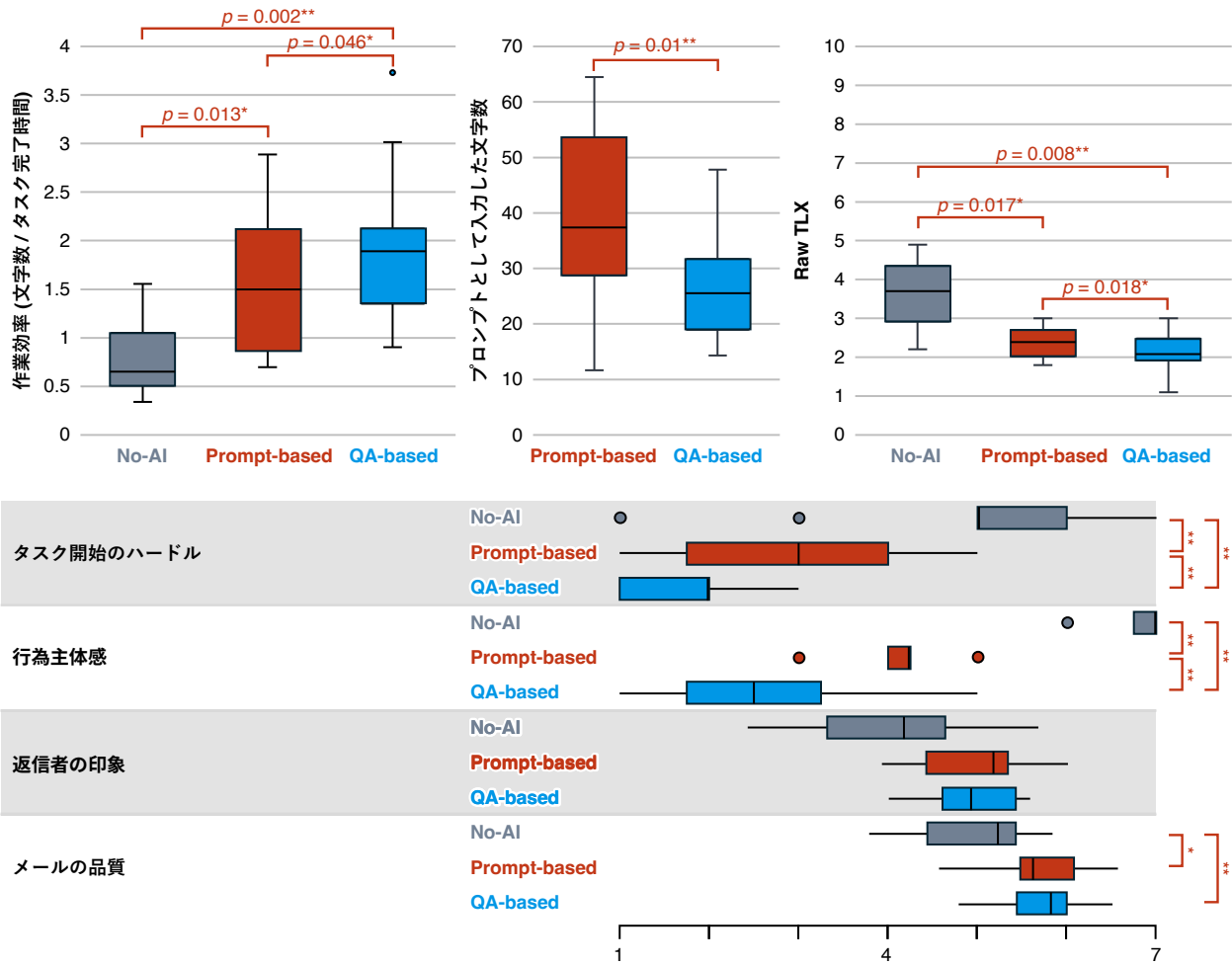


図 2. 各実験条件における定量評価指標の結果。下図において右の括弧は各条件間の有意差を表す: *($p < .05$), **($p < .01$).

や制御感、コミュニケーション相手に対して感じる心理的距離には悪影響を及ぼす可能性があり、将来的にシステムがユーザに受け入れられない可能性もある。したがって本研究では、ResQ を多角的に評価するために、2 種類のユーザ実験を行った。

3 評価実験

まず、12 名の参加者が、介入のない No-AI 条件、返信案を得るためのプロンプトを全て自身で入力する必要がある Prompt-based 条件、ResQ を用いた QA-based 条件の 3 条件で、架空のメールを用いて返信タスクを行った。結果は図 2 に示す通り、QA-based アプローチはユーザのメール返信作業体験の改善に、貢献することが示唆された。また QA-based アプローチが及ぼす中長期的な影響と、実務での利用体験を定性的に評価するために、9 名の参加者を対象に 5 日間の実環境実験を行った。参加者全員は、QA-based アプローチによって全体的なメール返信体験が改善し、今後利用し続けたいと回答した一方

で、4 名の参加者は行為主体感が低下したことを報告した。また一部の参加者は、コミュニケーションの質や量が改善したことを報告し、それによって相手との心理的距離が縮まったように感じたとも報告した。

4 考察と結論

実験結果から、QA ベースのアプローチはユーザの認知負荷を軽減し、作業効率を向上させることが示された。これは AI による質問と回答の選択肢の提示によって、ユーザのメール内容の理解と、返信方針のシステムへの伝達を簡略化したためと考えられる。また QA ベースのアプローチは、ユーザの返信作業の障壁を下げることで相手への応答時間を短縮し、返信の質を高める一方で、ユーザの行為主体感を低下させる可能性があることも示された。したがって QA ベースのアプローチは、ビジネスやカスタマーサービスなど、迅速かつ高品質な返信が求められる場面では効果的だが、個人的なやり取りでは別の介入を検討する必要がある。

参考文献

- [1] Anthropic. Claude: Next-Generation AI Assistant. Retrieved in July 10, 2024 from <https://www.anthropic.com/claude>, 2024.
- [2] A. Bastola, H. Wang, J. Hembree, P. Yadav, Z. Gong, E. Dixon, A. Razi, and N. McNeese. LLM-based Smart Reply (LSR): Enhancing Collaborative Performance with ChatGPT-mediated Smart Reply System, 2024.
- [3] V. Bellotti, N. Ducheneaut, M. Howard, I. Smith, and R. E. Grinter. Quality versus quantity: E-mail-centric task management and its relation with overload. *Human-Computer Interaction*, 20(1-2):89–138, 2005.
- [4] M. X. Chen, B. N. Lee, G. Bansal, Y. Cao, S. Zhang, J. Lu, J. Tsay, Y. Wang, A. M. Dai, Z. Chen, T. Sohn, and Y. Wu. Gmail Smart Compose: Real-Time Assisted Writing. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD '19*, p. 2287–2295, New York, NY, USA, 2019. Association for Computing Machinery.
- [5] L. Fu, B. Newman, M. Jakesch, and S. Kreps. Comparing Sentence-Level Suggestions to Message-Level Suggestions in AI-Mediated Communication. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, CHI '23*, New York, NY, USA, 2023. Association for Computing Machinery.
- [6] Y. Fu, S. Foell, X. Xu, and A. Hiniker. From Text to Self: Users' Perception of AIMC Tools on Interpersonal Communication and Self. In *Proceedings of the CHI Conference on Human Factors in Computing Systems, CHI '24*, New York, NY, USA, 2024. Association for Computing Machinery.
- [7] S. M. Goodman, E. Buehler, P. Clary, A. Coenen, A. Donsbach, T. N. Horne, M. Lahav, R. MacDonald, R. B. Michaels, A. Narayanan, M. Pushkarna, J. Riley, A. Santana, L. Shi, R. Sweeney, P. Weaver, A. Yuan, and M. R. Morris. LaMPost: Design and Evaluation of an AI-assisted Email Writing Prototype for Adults with Dyslexia. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS '22*, New York, NY, USA, 2022. Association for Computing Machinery.
- [8] Grammarly. Grammarly: Free AI Writing Assistance. Retrieved in July 10, 2024 from <https://www.grammarly.com>, 2024.
- [9] G. Mark, S. T. Iqbal, M. Czerwinski, P. Johns, A. Sano, and Y. Lutchyn. Email Duration, Batching and Self-interruption: Patterns of Email Use on Productivity and Stress. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16*, p. 1717–1728, New York, NY, USA, 2016. Association for Computing Machinery.
- [10] Microsoft. Microsoft Copilot: AI-Powered Assistance for Productivity. Retrieved in July 10, 2024 from <https://www.microsoft.com/en-us/microsoft-365/copilot>, 2024.
- [11] L. Nelson, R. Nairn, E. H. Chi, and G. Convertino. Mail2Tag: Augmenting email for sharing with implicit tag-based categorization. In *2011 International Conference on Collaboration Technologies and Systems (CTS)*, pp. 23–30, 2011.
- [12] OpenAI. ChatGPT (4o) [Large language model]. Retrieved in July 10, 2024 from <https://chatgpt.com/>, 2024.
- [13] S. Whittaker, V. Bellotti, and J. Gwizdka. Email in personal information management. *Commun. ACM*, 49(1):68–73, jan 2006.