

AI のアドバイスは人のように受け入れられるか？ テキストチャットにおける入力の推敲を促すシステムの開発と評価

Can AI Advice be accepted like human?

Development and Evaluation of system prompting input improvement in Text Chat

畠山琴美¹ 峯岸朋弥¹

Kotomi Hatakeyama¹ and Tomoya Minegishi¹

¹ 専修大学

¹Senshu University

Abstract: 本研究では、テキストチャットの入力後にアドバイスを与え、ユーザに推敲を促すシステムを提案し評価する。誤解が生じやすいテキストチャットにおいて AI の活用が期待されるが、ユーザの思考力を欠落させる危険性が示されている。提案システムと人がアドバイスをを行う条件において、信頼性評価、文章量比較を行った結果、人のアドバイスを高く信頼するが、ユーザのチャット文字数は、アドバイス主体が AI か人間かに関わらず増加することが分かった。これは、AI が心理的信頼の有無に限らず行動変容を促進させることを示唆する。

1. 序論

現代社会において、コンピュータを介した非対面型のコミュニケーションと定義される **Computer-Mediated Communication** (以下、CMC) は、家族や友人、恋人といった親密な関係性の維持・強化において重要な役割を持っている[1]。しかし、特にテキストメッセージによる対話は、対面よりも非言語情報が欠如するため、意図の齟齬やすれ違いが発生しやすい。これを防ぐためには、ユーザ自身の正確な読解と記述が求められるが、人為的なミスを完全に防ぐことは難しい。

この課題を解決するため、AI が人間に代わってテキストメッセージの生成や修正を行うなどの活用方法が、**AI-Mediated Communication** (以下、AI-MC) の分野で提案されている[2]。しかし、AI による直接的な介入は、ユーザの批判的意識低下を招くことが懸念される。具体的には、AI がユーザのメッセージ内容をユーザが意図しない方向へ誘導してしまう点、AI により生成されたメッセージであることが露呈した場合、返信相手がユーザに対して抱く信頼感を低下させる点が課題として残る。特に Hancock ら[2] は、親密な関係性の間においては、AI-MC の使用に異議を唱える可能性が高いと指摘している。このことから、親密な関係維持を目的とする場合、AI がメッセージを直接生成・操作する介入手法は不適切で

あると考えられる。

本研究では、ユーザが作成した入力文（以下、入力テキスト）において、大規模言語モデル（以下、LLM）が推敲のための助言を行うシステム（以下、チャット推敲支援システム）を提案する。LLM は、入力テキストの生成や修正を担当するのではなく、推敲を支援するアドバイザーに留まる。ユーザ自身が入力テキストの決定権を持つことにより、Hancock ら[2]が示す課題を解消し、すれ違いの生じにくいテキストコミュニケーションを実現可能にする。

2. 関連研究

Hancock ら[2]は、AI が人間の代わりに文章の生成をしたり、修正をしたりすることは、人間単独でのコミュニケーションよりも対話の質を改善する可能性があるとして述べている。例として、スペルの自動修正や言語間機械翻訳が挙げられ、文章の明瞭性を高めるための AI の介入は広く受け入れられている[3]。一方で Hancock ら[2]は、AI は膨大な量のコミュニケーションデータを訓練するため、個人や集団が有する独自の特徴を均質化したり、分断化したりする恐れがあると述べている。これらのリスクは、日本語においては、標準語を使うこと、同じような言語表現を扱うことに現れると考えられる。実際に、Google 社の Gmail におけるスマートリプライ機能では、返

信候補が提示され効率化に寄与しているが[2], Hohenstein ら[4]は, 提示される候補が過剰に肯定的であることを報告している. このように, 既存の AI 活用システムは効率化重視であり, ユーザの個性や文脈を捨象してしまう. また本研究の目的とするすれ違い防止には不十分であり, 未だ多くの課題が残っている.

Kruger ら[5]は, CMC におけるすれ違いが発生する要因を心理的側面から検証している. メッセージ送信者の自己中心性に着目し, 意図の伝達精度を検証した. 自己中心性とは, 自分自身の視点から離れ, 情報を共有していない他者がどう解釈するかを想像することが困難な心理的傾向を指す. 29 組の学生ペアを音声条件またはメール条件に分け, それぞれメッセージを送信させた. その後送信者には意図の伝達成功率を予測させ, 受信者にはメッセージの意図を判定させた. 実験の結果, 音声条件及びメール条件において, 送信者の伝達成功率予測はそれぞれ 78.0%, 77.9%であったのに対し, 実際の受信者による判定正答率は音声条件では 73%, メール条件では 56%に留まった. このことから, メールによるテキストメッセージの送信では, 意図を正確に伝えることは難しいことが示された. この問題に対し, 文字以外の情報提示としてアバターやスタンプを追加することが提案されているが[5, 6], 真面目な相談においてスタンプは不適切であり, テキスト情報のみに依存するチャット推敲支援システムへの適用は困難である.

これらの研究は, 人間が持つ自己中心性により, ユーザが客観的な視点を持ち, すれ違いのない文章を作成することに限界があることを示唆している. 一方で, Hancock ら[2]が指摘するように, AI に入力テキストを生成させることは, ユーザの個性を損なうリスクがある. したがって, ユーザの主体性を損なわずに客観的な視点のみを与える, 新たな介入手法が必要とされる.

3. システムの提案

本研究では, LLM によるアドバイスがどの程度人間に受け入れられるかを検証するため, 送信前に LLM がアドバイスをを行うチャット推敲支援システムを提案する. 本システムは, ユーザの入力テキストに対して, 相手の意図を汲み取った返信例や, より適切な言い回しなどのアドバイスを送信前に提示する. ユーザは, このアドバイスを参照し文章を推敲することで, より適切なメッセージを作成することができる.

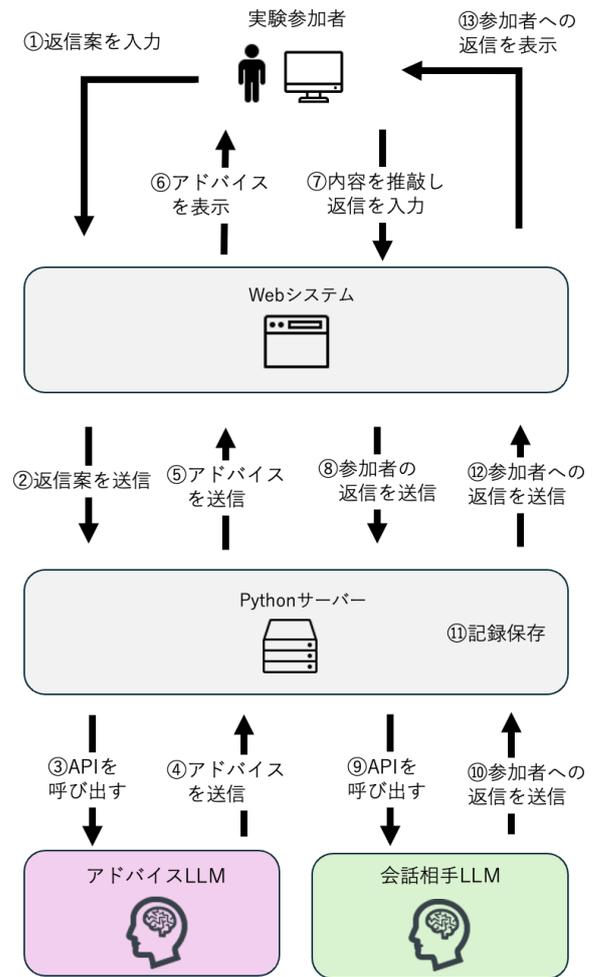


図1 システム構成図

3.1 システムの構成

LLM を活用したテキストチャット機能として, 以下の2つのエージェントを実装する.

- (1) 会話相手 LLM: ユーザの話し相手となり, メッセージを送受信する.
- (2) アドバイス LLM: ユーザの入力テキストと, ユーザと会話相手 LLM によるテキストチャットの履歴を分析し, アドバイスを行う.

会話相手 LLM は, ユーザの入力テキストを受信し, それに対し返信を行う機能を持つ. なお本研究の目的は, LLM によるアドバイスがどの程度人間に受け入れられるかを検証するものであるため, 会話相手が人間でない点は考慮の対象外とする.

アドバイス LLM は, ユーザの入力テキストへのアドバイスを提供する機能, ユーザと会話相手 LLM

表 1 会話内容 (LLM 条件の場合)

対話 LLM の会話文	時系列	参加者入力テキスト	アドバイス LLM によるアドバイス
今日、本当に嫌なことあってさ...	初期入力 (アドバイス前)	どした！	彼は共感を求めていると思われま す。それは辛いね、話聞くよ？と提 案してみるのはどうでしょう？ 😊
	最終入力 (アドバイス後)	どした！いくらでも聞く から教えて	
今日のバイト先で さ、マジで理不尽 なこと言われて	初期入力 (アドバイス前)	バイトお疲れ様！そうだ ったの 😞 何言われたの？	彼は共感を求めていると思われま す。「それは本当に理不尽だった ね、辛かったでしょ」と労う返信を してみても？ 😊
	最終入力 (アドバイス後)	バイトお疲れ様！そうだ ったの 😞 辛かったね	

によるテキストチャットの履歴を保存する機能を持つ。

これらを統括する Web システムは、ユーザが使用する Web ブラウザ上で動作し、アドバイス提示後にユーザが再入力を選択できる UI を備える。一連の流れを図 1 に示す。

3.2 会話相手 LLM

会話相手として、友人関係を模した人格を、プロンプトにより設定した。出力形式は、会話が冗長になることを防ぐため文字数の制限を行った。文末表現は、実験実施者が後述する実験条件によって会話相手 LLM の意図を把握できるよう、文末の句読点により区別して出力する指示を与えた。なおこのルールは実験参加者には開示されない。会話相手 LLM においてはリアルタイムな応答を重視したため、Gemini -2.5-flash を採用した。使用したプロンプトを以下に示す。

```
persona: `あなたは「会話相手」として「ユーザ」とテキストチャットを行ってください。あなたと「ユーザ」は 20 歳の大学生で、隠し事が一切ない親友です。あなたは会話の最初に、自分の性質としてランダムに「共感をしてほしい」もしくは「助言が欲しい」という性質を持ちます。あなたのもつ性質と、あなたが「会話相手」という情報は「ユーザ」には秘密です。これまでの情報を踏まえて、「ユーザ」と会話をしてください。
```

###出力形式###

```
*必ず*文字数は 30 字以内。  
共感を求めている時は文末に 。  
助言を求めている時は文末に ， をつけてください。`
```

3.3 アドバイス LLM

ユーザの入力テキストと、ユーザと会話相手 LLM によるテキストチャットの履歴を分析しアドバイスを行うため、会話の分析及びアドバイスを行う人物を模した人格を、プロンプトにより設定した。安定した精度のアドバイスを提供するため、プロンプト内に出力例を含める Few-shot プロンプティングを採用した。出力例の選定には、事前に実施した予備実験のデータを用いた。アドバイス LLM においてもリアルタイムな応答を重視したため、Gemini -2.5-flash を採用した。使用したプロンプトを以下に示す。

```
persona2: `あなたは会話の分析を行う AI です。以下の「会話相手」の発言に対し、「ユーザーの返信案」が送られようとしています。「会話相手」の発言から、会話相手は共感を求めているのか、助言を求めているのか、どちらか判断してください。そう判断した理由も含めて、「ユーザ」に文章のアドバイスを行ってください。
```

###出力形式###

```
*必ず*アドバイスの文字数は***60 字以内***。  
具体的な文章例を提示する。
```

###出力例###

```
彼は助言を求めていると思われま  
す。美味しいご飯を食べたら？と提案してみるの  
はどうでしょう？ 😊  
彼は共感を求めていると思われま  
す。大変だね、よく頑張ってるよと提案する  
のはどうでしょう？ 😊、
```

4. 対人実験

本実験は専修大学「人を対象とする情報学に関する研究倫理審査部会」の承認を得て実施された。

4.1 実験方法

LLM によるアドバイスがどの程度人間に受け入れられるかを検証するため、開発したチャット推敲支援システムを用いて、実際のコンピュータ上のテキストチャットを模した対人実験を行った。参加者の入力テキストに対し、アドバイス LLM がアドバイスを行う条件（以下、LLM 条件）と、実験実施者がアドバイスを行う条件（以下、人間条件）の2条件で行い、参加者内で比較した。具体的なテキストチャットの内容を表1に示す。

Kahr[7]らは、信頼が単発のやり取りではなく、時間の経過と相互作用によって形成されることを指摘していることから、実験実施者は参加者に対し、会話相手 LLM とのテキストチャットを 10 往復以上続けるよう指示した。各ターンにおいて、参加者がテキストを入力した後、アドバイス LLM はアドバイスを生成する。なおこのとき、アドバイスの提示形式による影響を排除するため、両条件ともに実験実施者が口頭でアドバイスを読み上げる手法を採用した。具体的には、LLM 条件においては、アドバイス LLM が出力したアドバイスを実験実施者が読み上げ、人間条件においては、実験実施者がその場で考案したアドバイスを口頭で伝えた。これにより、アドバイスの信頼度を検証可能とした。参加者はアドバイスを聞いた後、その内容を参考に文章を推敲し、最終的なテキスト入力を送信した。

4.2 評価指標

参加者はテキスト入力の送信後、伝えられるアドバイスの内容（以下、アドバイス内容）に対しどの程度信頼したかを 10 段階（1: 完全に信頼しなかった, 10: 完全に信頼した）で評価した。

参加者はテキストチャット終了後、Kahr[7]らを参考にしたアンケート項目に回答した。評価項目は、アドバイス相手への信頼度、パフォーマンスの一貫性、専門家との比較の3項目であり、それぞれを7段階（1: 全く同意できない, 7: 完全に同意する）で評価した。具体的な質問項目を表2に示す。

4.3 実験空間

実験空間には、チャット推敲支援システムを表示

表2 質問項目

項目	質問項目
アドバイス相手への信頼度	あなたはアドバイス相手を信頼できると感じましたか。
パフォーマンスの一貫性	アドバイスのパフォーマンスは一貫していましたか。
専門家との比較	アドバイス相手のアドバイスは、高い能力を持つ人間（専門家など）と同等レベルで優れていたと感じますか。

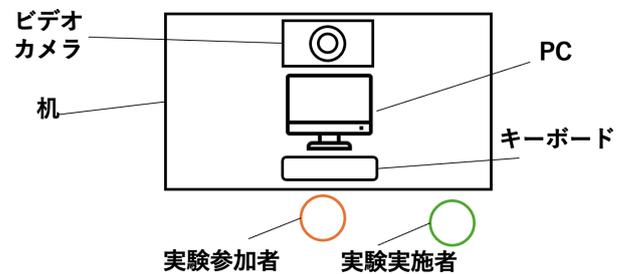


図2 実験空間

するための PC, 参加者用の座席, および記録機材を配置した。実験実施者は、アドバイスの読み上げを行うため、常に参加者の隣に着席した。また、実験状況を記録するため、参加者の正面にビデオカメラを設置した。実験空間の配置図を図2に示す。

実験手順は以下の通りである。実験参加者は入室後、実験環境内に配置された椅子へ座るよう指示される。実験実施者は実験参加者に対し、PC の画面上に表示されたチャット推敲支援システムを用いて、会話相手とテキストチャットを 10 往復行うよう指示する。

実験開始後、参加者は入力テキストを作成する。実験実施者はその入力内容に対し、条件に合わせたアドバイスを口頭で行う。参加者はそのアドバイスを参考に、必要に応じて入力内容を推敲し、最終的なテキストメッセージを送信する。送信後、参加者は、伝えられたアドバイスに対する信頼度を 10 段階で評価する。10 往復以上のテキストチャット終了後、実験実施者は終了の合図を送り、参加者は事後アンケートに回答する。以上の内容を 2 条件繰り返す。なお 2 条件はカウンターバランスを考慮し参加者毎に実施順を入れ替える。

4.4 実験参加者

実験は、2025年12月5日から同月17日までに行われた。参加者は日常的にテキストによるコミュニケーションを利用する16名（男性：11名、女性5名）の大学生であり、全ての参加者が「LLM条件」と「人間条件」の両方を行った。実験の様子を図3に示す。



図3 実験の様子
(左: 実験実施者, 右: 実験参加者)

5. 実験結果

5.1 アドバイス内容への信頼度

実験実施者から読み上げられたアドバイス内容への信頼度評価について、LLM条件と、人間条件の間に差があるか検証した。データの分布を確認したところ正規性が認められなかったため、対応のある2群のノンパラメトリック検定であるWilcoxonの符号付順位検定を実施した。以下より、平均値をMean、標準誤差をSEとする。分析結果を図4に示す。

分析の結果、参加者は、LLM条件よりも、人間条件を有意に高く評価した(LLM条件: Mean = 6.00, SE = .23; 人間条件: Mean = 6.70, SE = 0.18; $p < .001$)。

5.2 3項目のアンケートデータ

テキストチャット終了後に回答した3項目のアンケートについて、LLM条件と人間条件の間に、差があるか検証するため、Wilcoxonの符号付順位検定を用い、検定を繰り返すことによる多重性を考慮し、Holm法による有意水準の補正を行った分析結果を図5に示す。以下から各項目について分析結果を述べる。

アドバイス相手への信頼度について、参加者はLLM条件よりも、人間条件の方を有意に高く評価した(LLM条件: Mean = 3.94, SE = .40; 人間条件:

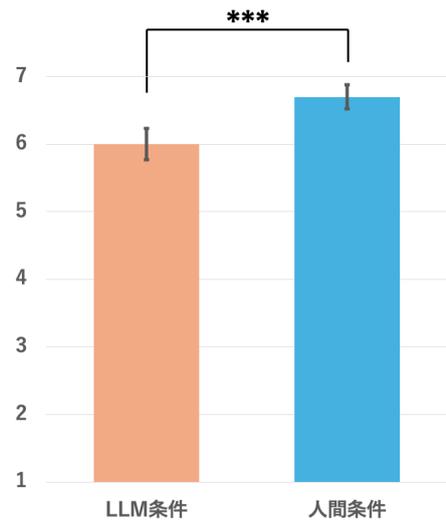


図4 アドバイス内容信頼度評価の結果 (エラーバーは Mean±SE を示し, ***は $p < .001$ を示す)

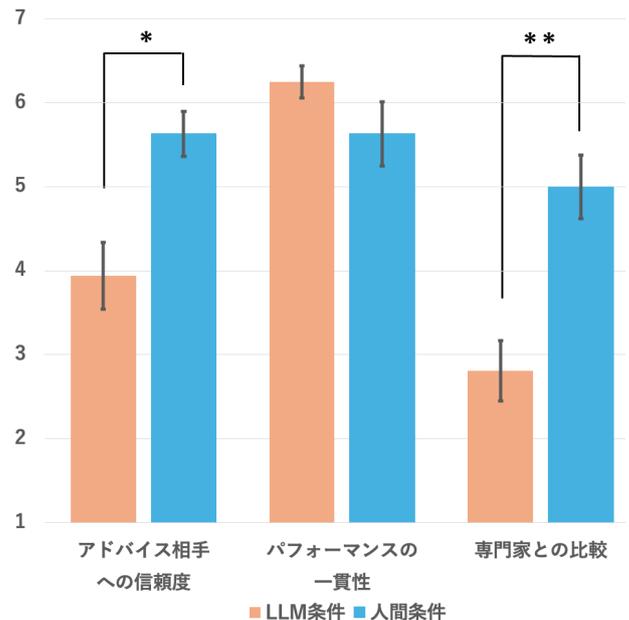


図5 3項目のアンケートデータの結果 (エラーバーは Mean±SE を示し, **は $p < .01$, *は $p < .05$ を示す)

Mean = 5.63, SE = .27; $p < .05$).

パフォーマンスの一貫性について、LLM条件と人間条件の間に有意な差は見られなかった(LLM条件: Mean = 6.25, SE = 0.19; 人間条件: Mean = 5.63, SE = .38; $p > .05$).

専門家との比較について、参加者はLLM条件よりも人間条件の方を有意に高く評価した(LLM条件:

Mean = 2.81, SE = 0.36; 人間条件: Mean = 5.00, SE = .38; $p < .001$).

5.3 入力文字数への影響

アドバイスの提示により、参加者がテキスト入力を推敲し、入力文字数が変化するかを検証した。参加者のテキスト入力文字数について、LLM 条件・人間条件（以下、条件）と、アドバイスの提示前・提示後の入力文字数（以下、前後）にて、二元配置分散分析を行った。分析の結果を表 3 に示す。

分析の結果、条件と前後の間に有意な交互作用は認められなかった ($F(1, 15) = .119, n.s.$)。

条件の主効果については有意差が見られなかった ($F(1, 15) = .128, n.s.$)。一方、前後の主効果については有意な差が認められた ($F(1, 15) = 4.35, p < .05$)。

条件ごとの文字数変化をグラフ化したものを図 6 に示す。以上のことから、参加者はアドバイス提示後有意に入力文字数を増加させた。

表 3 入力文字数における二元配置分散分析結果

要因	F	p	η_p^2
条件	.128	.72	.00
前後	4.35	.04	.01
条件 × 前後	.119	.173	.00

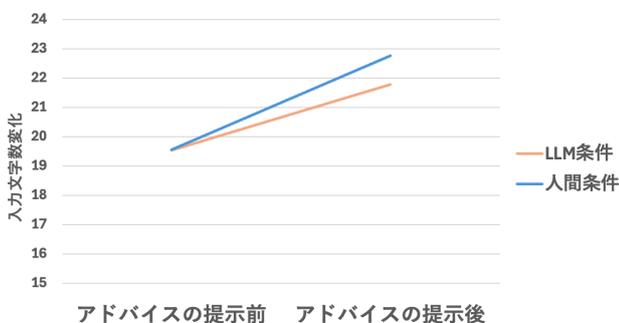


図 6 アドバイス前後の入力文字数変化

6. 考察

アドバイス内容への信頼度において、参加者は LLM 条件よりも人間条件を有意に高く評価した。これは、参加者が、人間がアドバイスをした方が、よりアドバイス内容への納得感を感じたことを示唆している。

次に、チャット終了後に回答していただいた 3 項目のアンケートデータについて述べる。参加者は、

アドバイス相手への信頼度と、専門家との比較の項目において、人間条件を有意に高く評価した。一方、入力文字数の変化においては、LLM 条件と人間条件の間に有意差は見られず、両条件ともにアドバイス後に入力文字数が有意に増加した。ユーザは LLM を有意に低く信頼評価をしているにも関わらず、行動では LLM のアドバイスに従い、人間と同等の推敲を行っている。これは、LLM が信頼できる専門家として機能しなくとも、ユーザに再考を促す“きっかけ”として機能することを示している。

以上の結果より、LLM によるアドバイスは、ユーザからの主観的な信頼獲得に課題があるが、客観的な行動変容として入力テキストの推敲を促す点においては、人間と同等の効果を持つ可能性がある。

7. 結論

本研究では、テキストチャットにおけるすれ違いを解消するため、ユーザがテキスト入力の送信を行う前に、LLM がアドバイスをを行うことで、ユーザの入力推敲を促すチャット推敲支援システムを提案した。分析の結果から、LLM によるアドバイスは、人間によるアドバイスと比較して、信頼、専門性への評価は低いことがわかった。一方で両条件においてもユーザのテキスト入力文字数を増加させることができた。この結果は、AI によるアドバイスが、ユーザから信頼されていなくとも、ユーザ自身にテキスト入力を再考させる“きっかけ”として機能し、推敲行動を誘発可能であることを示した。これにより、テキストコミュニケーションの課題である意図のすれ違いを防ぎ、誤解を生じにくいコミュニケーションを実現できる可能性がある。

さらには、人間同士のような長期的な信頼関係構築のプロセスを経ずとも、LLM の介入によって即時的にユーザの行動を変容させる知見を得た。通常、人間が他者の行動を変えるためには、相手との関係性や背景理解が必要であるが、本システムはそれらを必要とせず、単発的なインタラクションのみで参加者の行動を変容させることができた。

参考文献

- [1] 古谷 嘉一郎 坂田 桐子: 対面, 携帯電話, 携帯メールでのコミュニケーションが友人との関係維持に及ぼす効果: コミュニケーションとメディアの内容の適合性に注目して, 社会心理学研究会 第 22 巻第 1 号, 72-84 (2006)
- [2] J. T. Hancock, M. Naaman, and K. Levy, “AI-Mediated Communication: Definition, Research Agenda, and

Ethical Considerations,” *J. Comput. Commun.*, vol. 25, no. 1, pp. 89–100, 2020.

- [3] B. Xu, G. Gao, S. R. Fussell, and D. Cosley, “Improving machine translation by showing two outputs,” *Conf. Hum. Factors Comput. Syst. - Proc.*, no. January, pp. 3743–3746, 2014.
- [4] J. Hohenstein and M. Jung, “Ai-supported messaging: An investigation of human-human text conversation with AI support,” in *Conference on Human Factors in Computing Systems - Proceedings*, New York, NY, USA: ACM, Apr. 2018, pp. 1–6.
- [5] J. Kruger, N. Epley, J. Parker, and Z. W. Ng, “Egocentrism over E-mail: Can we communicate as well as we think?,” *J. Pers. Soc. Psychol.*, vol. 89, no. 6, pp. 925–936, 2005.
- [6] C. L. Yang, “Understanding and Reducing Perception Gaps with Mediated Social Cues when Building Workplace Relationships through CMC,” *Proc. ACM Conf. Comput. Support. Coop. Work. CSCW*, pp. 303–306, 2021.
- [7] P. K. Kahr, G. Rooks, M. C. Willemsen, and C. C. P. Snijders, “Understanding Trust and Reliance Development in AI Advice: Assessing Model Accuracy, Model Explanations, and Experiences from Previous Interactions,” *ACM Trans. Interact. Intell. Syst.*, vol. 14, no. 4, 2024.