Will AI Console Me when I Lose my Pet? Understanding Perceptions of AI-Mediated Email Writing

YIHE LIU, Georgia Institute of Technology, USA

ANUSHK MITTAL, Georgia Institute of Technology, USA

DIYI YANG, Georgia Institute of Technology, USA

AMY BRUCKMAN, Georgia Institute of Technology, USA

Large language models are increasingly mediating, modifying, and even generating messages for users, but the receivers of these messages may not be aware of the involvement of AI. To examine this emerging direction of AI-Mediated Communication (AI-MC), we investigate people's perceptions of AI written messages. We analyze how such perceptions change in accordance with the interpersonal emphasis of a given message. We conducted both large-scale surveys and in-depth interviews to investigate how a diverse set of factors influence people's perceived trust in AI-mediated writing of emails. We found that people's trust in email writers decreased when they were told that AI was involved in the writing process. Surprisingly trust increased when AI was used for writing more interpersonal emails (as opposed to more transactional ones). Our study provides insights regarding how people perceive AI-MC and has practical design implications on building AI-based products to aid human interlocutors in communication.

CCS Concepts: • **Human-centered computing** → **Interaction paradigms**.

Additional Key Words and Phrases: Artificial Intelligence-Mediated Communication (AI-MC); Computer-Mediated Communication (CMC); Artificial Intelligence; Trust

ACM Reference Format:
Yihe Liu, Anushk Mittal, Diyi Yang, and Amy Bruckman. 2022. Will AI Console Me when I Lose my Pet? Understanding Perceptions of AI-Mediated Email Writing. In *CHI Conference on Human Factors in Computing Systems (CHI '22), April 29-May 5, 2022, New Orleans, LA, USA.* ACM, New York, NY, USA, 20 pages. https://doi.org/10.1145/3491102.3517731

"If I was told that it was AI-written, I would not be happy about it. If it just popped up in my inbox, and I don't know that it is AI-written, then I would be like, 'Yeah, this is a good email' because all of them were actually good emails." (Participant 4)

1 INTRODUCTION

Since the advent of computers, communication has been increasingly mediated by technology. Computer-mediated communication (CMC) has been comprehensively studied as an established field of research to understand its influence on our communication [3, 22, 29]. While face-to-face communication affords social cues like body language that facilitate our perception of the communicator, CMC often limits such cues. The Hyperpersonal Model of CMC [56] explains that receivers often over-interpret cues from senders because of reduced availability of cues in computer-mediated communication. Such over-interpretation occurs in our perceptions of personality traits [38–40], social identities [32], cognitive styles [44], and trustworthiness of senders [33].

We are now entering a new paradigm of communication in which intelligent agents operate on behalf of a communicator by modifying, augmenting, or even generating messages to accomplish communication goals [17]. It is no longer a mere theoretical exploration as language models are increasingly achieving human-level performance in mediating communication [7, 45, 53]. Applications of AI models are already shaping the way we communicate today [24, 52, 60], presenting serious implications for people's perception of such AI-mediated communication (AI-MC).

Algorithmic "Smart Reply" suggestions increase communication efficiency, the use of positive emotional language, and positive evaluations by communication partners [24, 25]. However, people are evaluated more negatively if they are suspected to be using algorithmic responses [24]. Despite its many benefits, AI mediation risks changing users' language production and continues to be viewed negatively. However, this suspicion of algorithmic responses might not hold true as language models become increasingly human-like.

Since the launch of the GPT-3 API in mid-2020 [1], AI-MC tools have been rapidly deployed by millions of people around the world. These include tools that convert short bullet points into long form e-mail messages[2], that write advertisement copy based on short product descriptions [3], that write compelling email subject lines to optimize the chance a message will be opened [4] and much more. In fact, this introduction is augmented using multiple AI systems. The smart autocomplete suggestions provided by Google Docs completed various sentences after typing mere two or three words. *ShortlyAI* [5] can write complete paragraphs based on the title of the paper and a short abstract. The contents of this introduction were inspired by several of the suggestions by their AI system. While we only use those AI generated texts as mere suggestions, one could imagine using them as-is, without needing any edits.

This raises the question, does it make you more sceptical or less trusting about this introduction as you realize that it was augmented using an AI system? What if you were not privy to the AI mediation we mentioned above? Answering these hard questions is the purpose of this paper. There is a dearth of literature and empirical evidence to understand the use of AI-MC [17]. We extend current literature by posing our research question: **How does the involvement of AI in writing emails affect users' perceived trust in the communication?** We conducted a mixed-methods study including an online survey and 1:1 interviews. Participants were presented with email messages that were actually human-written but were told to be either human-written or mediated through an AI system. This allowed us to remove varying quality of generated text [46] as a possible cofactor for human evaluations. We presented participants with emails of varying interpersonal emphasis to further understand any interaction effects.

We found that participants are generally less trusting of emails perceived to be mediated through an AI. Furthermore, we found that as interpersonal emphasis increases perceived trustworthiness increases and this trend directionally holds true for all degrees of AI mediation. This contradicts the warranting theory [13] that suggests AI-augmented communication is perceived as more warranted if it is objective (low interpersonal emphasis) and Jakesch et al. [28] who suggest that AI mediation in self-descriptions (high interpersonal emphasis) should activate concerns about deception. During our interviews, we found that participants were indifferent to the presence of an AI system when rationalizing their perception. They were absorbed by the human-like quality of presented messages which lead them to focus on paralinguistic cues to assess given communication as predicted by the CMC literature [55, 56]. However, when asked about the use of such AI systems, they overwhelmingly rejected it for highly interpersonal messages, expressing the

[1] https://openai.com/blog/openai-api

[2] https://techcrunch.com/2020/11/12/othersideai- raises- 2- 6m- to- let- gpt- 3- write- your- emails- for- you

[3] https://techcrunch.com/2021/03/17/gpt-3-powered-copy-ai-raises-2-9m-in-a-round-led-by-craft-ventures

[4] https://www.conversion.ai

[5] https://shortlyai.com/

content as *"fake"* and the sender as a *"traitor"*. This leads to interesting ethical and moral implications for UX designers of such systems.

The contributions of this paper are as follows:

- **Contribution 1:** Replication of previous work in a novel context (email). Previous work studied perceived trustworthiness of "smart replies" [24] or text messages [51] that are short and sometimes follow man-made rules. As language models become more competent at writing and more widely used for writing, it becomes essential to study the impact of long-form writing such as emails. Email is also a widely used communication medium and its content can have many topics and domains. This enables better generalizability of our results than those of single-domain communication systems like Airbnb [28].
- **Contribution 2:** Investigation of the interaction between interpersonal emphasis and the effect of AI-mediated communication on trust. Our work shows that interpersonal emphasis is positively correlated with perceived trustworthiness regardless of whether or not the message is written by AI. This contradicts with previous studies that explores the interaction [13]. We provide possible explanations for this contradiction and suggest scenarios where our findings apply, and discuss the significance of our results in designing AI-writing tools.
- **Contribution 3:** Complementing prior work by a qualitative analysis of in-depth interviews. Prior work suggests perceived AI involvement decreases trust. We received nuanced explanations of this effect via interviews with 10 participants. We discussed the broader implication of the participants' response on the use of AI-mediated communication for social issues.

2 RELATED WORK

2.1 AI-mediated Communication

Our study is motivated by the increased capacity of AI systems to mediate communication between people by operating on their behalf. Today's AI systems can modify, augment, and generate messages to accomplish interpersonal goals. Such augmentation has traditionally been studied under the field of computer mediated communication (CMC). December [14] operationally defines CMC as "*a process of human communication via computers, situated in particular contexts, engaging in processes to shape media for a variety of purposes.*" This understanding of CMC has meant studying the use of technology tools like spell-check and SMS for communication.

Analyzing the existing literature in CMC, Hancock et al. [17] acknowledge that the literature does not account for the extremity of dimensions such as the extent, the goals, and the autonomy with which technology can optimize messages. They term this AI-mediation of communication as AI-MC and propose a research agenda to build foundational empirical understanding about the designs and implications of these technologies. Our paper responds to their call for action in understanding people's perception of the use of such AI systems. We also provide insights into possible design implications to build AI systems that help humanity communicate.

Advancing the field of AI-MC, Hohenstein et al. [24] study the use of algorithmic responses or "smart replies" that allow senders to choose short replies from recommendations based on the ongoing conversation. They ask participants to engage in instant messaging with one another either using smart replies or not. They find that senders prefer the use of algorithmic responses, and receivers find a sender less cooperative if they suspect the usage of AI. On a different note, Hohenstein and Jung [26] suggest that the possible presence of AI in communication can lessen the responsibility assigned to the human sender when interactions are unsuccessful. These studies leave a gap in the AI-MC literature on the impact of: (1) perceived autonomy of AI (2) ambiguity about the use of AI by sender (3) familiarity of the

AI system by the receiver (4) long-form writing. To address the issue of perceived autonomy of AI, our study uses all human-written messages but tells the participants that some were written by AI [28, 51]. We also clearly define if a message has been augmented by AI, removing any ambiguity about its use by the sender. We randomly assign participants to different levels of perceived AI autonomy to compare perceptions between subjects. Unlike Jakesch et al., we do not familiarize participants with the AI system referenced to be used by the sender.

Finally, Calderwood et al. [9] explore how fiction writers use generative language models during their writing process. They find that novelists interact with these models to generate ideas for describing scenes and characters, to create antagonistic suggestions, and for other descriptive work. Long-form writing through increased AI autonomy is becoming increasingly mainstream. The impact of such long-form writing in AI-MC is yet to be seen. We choose emails as the choice of our communication medium as they are widely used and often include such long-form writing for communication as compared to smart suggestions. [47] ask participants to pose as senders and assess email reply suggestions across different scenarios. They find that contextual factors like social ties and the presence of salutations impacts sender's perceptions of email correspondence. We aim to extend this study by understanding how receivers perceive such email correspondences told to be mediated by AI.

2.2 Trustworthiness

We choose trust to study human perception as it is among the most fundamental aspects of human communication. Trust forms the root of human relationship and is reflected in collaborative behaviors such as willingness to depend, give information, and make purchases [42]. In CMC, trust and deception in online self-representation have been studied extensively. For instance, Hancock et al. [19] suggests that deception about dating profiles is common, but that the magnitude of the deceptions is usually small. [59] find the frequency of deception varies with different media. [18] suggests that, regardless of the medium, warrants like the use of real name or photo can suppress the frequency and seriousness of deception.

With the emergence of AI-MC from CMC, it is important to ensure that trust is not lost due to the involvement of an AI system. The Hyperpersonal Model [55, 56] suggests that compared to ordinary face-to-face situations, a computer-mediated communication allows a sender a greater ability to strategically edit self-presentation, enabling an optimized presentation of one's self to others. This suggests that if the AI can aid in improving such self-presentation, it may also increase a receiver's trust of the sender. The Hyperpersonal Model also suggests that receivers value linguistic cues in their perception of a communication. The structure and linguistic cues of e-mail messages might be a relevant factor for participants to judge trustworthiness.

Building on previous work, We define trustworthiness as an attribute of the trustee (or the sender) [20, 30] and trust as exhibited by a trustor (or the receiver) [4, 10]. Previous research have established the key ingredients in creating trust as a list of highly correlated concepts [31, 58]. For our study, we follow [41] that defines a trustworthiness scale based on three dimensions: ability, benevolence, and integrity. Ability reflects the group of skills, competencies, and characteristics that allow a party to have influence within some domain. Benevolence is the extent to which a trustee is believed to want to do good to the trustor. Finally, integrity is defined as the trustor's perception that the trustee adheres to a set of principles that the trustor finds acceptable. We designed a survey question for each dimension of trust to quantify the dimension with a 1-5 likert scale. We averaged participants' answers to these three questions to get a trust index following past studies [37].

In other work, trust has been defined as an individual characteristic, similar to a personality trait, rooted in life experiences and societal norms [2, 5, 48]. A propensity to trust others [16] might influence a receiver's trust of AI mediation in communication. Similarly, attitudes towards computers and AI might influence the receiver [1]. Our study acknowledges these personality traits as possible co-factors that can influence perceived trustworthiness in AI-MC and we study them based on shortened scales evaluated by [30, 43, 50].

3 METHODS

Before talking about our study design, we first define key terms:

- **AI condition:** the perceived degree of AI's involvement in writing. Participants see one of the three AI conditions in our experiments: 1. the email is written by human. 2. the email is written by human with the help of AI. 3. the email is written by AI.
- **Interpersonal Emphasis:** the perceived degree of how much the content of the email involves relations between persons.
- **Subject Expertise:** the familiarity of a participant on the topic of an email
- **AI attitude:** an index of the participant's positive and negative attitudes toward AI, as defined by [50].
- **Computer attitude:** an index of the participant's positive and negative attitudes toward computers overall, following [43].
- **Propensity to Trust:** a measure of how much the participant trusts others in general [30].

In this study, we ask *how does the involvement of AI in writing emails affect users' perceived trust in the communication?* Additionally, our study is based on the following questions:

- **RQ1:** Does AI condition affect trustworthiness?
- **RQ2:** Does Interpersonal Emphasis affect trustworthiness under different AI conditions?
- **RQ3:** Do subject expertise, AI attitude, and computer attitude impact trustworthiness?

Our study has two parts: a quantitative part and a qualitative part. First, in our quantitative experimental setup, participants read 12 email messages. This number of emails is used to set the average survey response time at around 20 minutes. Participants were asked to assume to be receivers of those messages. We used an online survey to allow participants to view these email messages and provide their trustworthiness ratings. After analyzing our quantitative data, we developed qualitative questions to understand some key aspects of the survey study's results. The 10 interview participants first took the online survey and then answered questions about their thought processes. We used the "*Wizard- of-Oz*" approach as used by various other studies of interpersonal communication [12, 15, 28, 35] (i.e. we provided only human-generated communication messages under the guise of AI usage changes). We chose to use human-written messages to remove the quality of the text [21] as a variable as much as possible because our goal was to gain insight into how people feel about texts written by AI. Furthermore, we wish to study the future potential impact of AI-MC rather than its current capabilities. It is clear that the ability of AI systems to generate human-like messages is continuously increasing [7].

We formed our hypothesis as follows:

- **H1:** Messages under the complete AI agency condition would rate low on trustworthiness as compared to those with no priming and complete human agency.
- **H2:** Under the complete AI agency condition, emails with higher Interpersonal Emphasis levels would show lower perceived trustworthiness score. Moreover, this effect would be reversed for emails in the complete human agency condition.

Scenarios	**Example Emails**
Scenario Product Inquiry	Hi, I was looking at your website and wondering if it's possible to grow tomatoes inside in the winter in an apartment? Is there a particular type of tomato that would be easier to grow? Would I need a grow light? Looking forward to your reply. Thank you, Drew
Scenario Party Invitation	Hi Jules, How are you doing? It's been quite a while since we met each other. 8 months to be exact! I'm super excited to host my birthday party this weekend and wanted to check in with you, if you might be able to make it. We plan to start around 9pm, but you can come by a bit late if that works better. I know you get off work late on weekends! Please bring you brother as well if he's free. Always fun to see him.. Looking forward to seeing you! Best, Bailey
Scenario Consolation of Pet Loss	Hi Elliott, Dave just let me know about Spot dieing. I'm so sorry for your loss. I know words wouldn't be able to comfort you at this time but I just wanted to write in to let you know I'm thinking about you. I'm here if I can help in any way. I cherish our memories with Spot together. The way he'd fetch ball with Jimmy!! I'll share our album with you on Google Photos. Take care! Quinn

Table 1. Example emails for each Interpersonal Emphasis Scenario.

- **H3:** AI attitude, computer attitude, and general propensity to trust would be positive correlated to perceived trustworthiness.
- **H4:** Subject expertise would positively impact perceived trustworthiness.

Independent Variables: While all participants saw the same emails, they each received one of the four AI priming conditions: no priming (control), complete human agency, shared human & AI agency, complete AI agency. Each participant encountered emails of all three Interpersonal Emphasis levels: low, medium, and high.

Dependent Variables: The three-item trustworthiness scale addressed the three dimensions of trust: Ability, Benevolence, and Integrity [41]. We recorded participants' perceived trustworthiness of the emails – measured as a cumulative score of the three-item trustworthiness scale.

3.1 Survey Study Design

The survey consisted of (a) a consent form (b) 12 email messages and questions pertaining to them and (c) a closing survey. The consent form informed participants that the purpose of this study was to examine the participant's understanding of written textual communication. We let them know that we would not inform them about everything within the study in advance and that, at the end of the survey, they would be debriefed and given the option to have their data discarded. Participants were then asked to answer optional demographic questions (i.e. their age, gender and ethnicity). We collected this information to provide a demographic composition of our participants in order to better understand the results.

Then, participants were presented with 12 emails in a randomized order. According to the AI condition, a relevant prompt was prepended to each email presented to the participants. For the complete human agency condition, we used the message "*The following email was written by sender.*" For the shared human and AI agency condition, we used the message "*The following email was written by sender with the help of a smart auto-complete system*", and finally, for the extreme AI agency, we used the message "*The following email was written by an advanced AI system on behalf of sender*". Here, sender reflected the name of a sender picked from the list of most popular unisex names in the US. This was done to avoid any biases that might be associated with names and subjects of communication and/or the participants.

We crafted 4 emails with different topics for each of the three categories of Interpersonal Emphasis (i.e. low, medium and high). Table 1 shows an example email for each Interpersonal Emphasis Scenario.

- **Scenario Product Inquiry:** the email scenario was about a person inquiring about customer support for a given product. This email was used to reflect something boring and technical that carried no interpersonal connotation (low emphasis).
- **Scenario Party Invitation:** the email scenario was about a person writing an email to a friend inviting them to a uniquely planned party. This was a party invitation and it had more of an Interpersonal Emphasis for the receiver (medium emphasis).
- **Scenario Consolation of Pet Loss:** the scenario is about a person emailing to comfort a friend who just suffered the loss of their pet (high emphasis).

For each email, the participants were prompted with "*Please answer the following questions based on the email you just read. Please make sure to read each question carefully*" with the three-item trustworthiness scale we developed for each of the scenarios following the three dimensions of trust described by Mayer et al. [41]. The responses to the three-item trustworthiness scale were recorded on a five-point Likert scale labelled as "Strongly disagree," "Somewhat disagree," "Neither agree nor disagree." "Somewhat agree," and "Strongly agree." We then asked each participant a question measuring their subject expertise since knowledge about a particular topic or scenario might bias their perception of trustworthiness of the message if the presented viewpoint did not match their own.

For Scenario Product Inquiry, the questions asked were:

- **Ability:** Do you expect the customer to buy this product?
- **Benevolence:** Is the customer concerned that they ask all adequate information from customer support?
- **Integrity:** Is the customer hiding any information that they already know?
- **Subject Expertise:** Are you familiar with the product being referred to in this email by the customer?

For Scenario Party Invitation, the questions asked were:

- **Ability:** Do you think the sender is capable of hosting this party?
- **Benevolence:** Do you believe that the sender will actually hold this party?
- **Integrity:** Do you think the sender is hiding any information?
- **Subject Expertise:** Have you ever received an email invitation (or sent one) for a special celebration party?

For Scenario Consolation of Pet Loss, the questions asked were:

- **Ability:** Do you believe that the sender actually understands the loss of their friend?
- **Benevolence:** Do you believe that the sender is actually concerned for their friend?
- **Integrity:** Do you think the sender actually believes in what they says?
- **Subject Expertise:** Have you ever (or has a close friend or family member) experienced the loss of a pet?

Finally, for the closing survey, we ask participants to rate the following statements measured on a five-point Likert scale with labels "*Strongly disagree,*" "*Somewhat disagree,*" "*Neither agree nor disagree,*" "*Somewhat agree,*" "*Strongly agree.*" The order of these questions was randomized.

- (a) one-item disposition to trust scale [37] based on "*Most people can be trusted*"

- (b) shortened two-item Computer Attitude Scale [43] based on "*Computers can eliminate a lot of tedious work for people.*" and "*Computers are lessening the importance of too many jobs now done by humans.*" (reverse coded)
- (c) shortened two-item AI Attitude Scale [50] based on '*I am interested in using artificially intelligent systems in my daily life*" and "*People like me will suffer if Artificial Intelligence is used more and more*" (reverse coded).

Once participants answered all questions, they were presented with the debriefing document that revealed the true purpose of the study along with the AI priming condition to which they had been assigned. At this point, they had the right to have their data dismissed with no compensation-based penalties.

3.2 Interview Study Design

1. Survey (30 minutes). Participants completed the survey independently without spoken or visual communication with the interviewer.

2. Interview (25 minutes). The questions were asked in a semi-structured format in which we followed up on what the participant said. The following bullet points comprise the key topics we aimed to explore and examples of the types of questions that were asked. For each question and response:

- On<question x> you answered <response>:
  - Why did you choose that response? Can you walk me through your reasoning?
  - How familiar are you with the subject discussed in this email? Did that shape your response?
  - How would you describe how the email is written?
- We asked you about your opinion of AI. Tell me about that. Why did you choose <response>? Can you walk me through your reasoning?
- Would you find it helpful to have AI tools to write emails? If yes, under what circumstances would you find it helpful?
- Are AI writing tools ever inappropriate? If so, under what circumstances?

3. Debrief and Conclude(5 minutes). We performed debriefing with our interviewees by explaining the Wizard-of-Oz approach and the goal of our study.

In the end, we collected 223 minutes of interviews. Interviews were professionally transcribed. The transcripts were then analyzed using MaxQDA, a software tool that enables easy sorting, structuring, and categorizing of large amounts of qualitative data. We used MaxQDA because it simply speeds up the qualitative evaluation process without suggesting interpretations. We categorized the quotes into groups according to the Research Questions that they answer.

As researcher self disclosure is an important part of qualitative research, we note that one member of the research team is strongly optimistic about the potential of AI, and one is strongly pessimistic. The other two team members are in the middle.

3.3 Ethical Considerations

At the end of the experiment, we informed participants the method and goal of our study. We did not collect name or other personal information, and therefore, the identity of the participant remains unknown to the research team. In addition, we only gathered data relevant to performing our analysis, such as those relating to the dependent or independent variables. To provide fair compensation to our participants, Mturkers were offered an equivalent of United

States federal minimum wage in terms of the time expected to complete the survey. Interview participants were offered $20 for one hour of video meeting.

4 RESULTS

4.1 Participants

We recruited 229 participants via Amazon Mechanical Turk (MTurk). Surveys were administered via Qualtrics. We required all participants to have approval ratings greater than 90 percent and to be Master-Mturkers. According to Amazon, a Master-Mturker is someone who has consistently demonstrated a high degree of success in performing a wide range of HITs across a large number of tasks. We did not limit participation to people in the US and the inclusion of non-US participants may affect the result. Finally, we excluded participants who failed the following two attention checks. First, participants needed to summarize the email messages in their own words. We excluded participants whose summarization was directly copied from or contains incorrect statements about the emails' content. Second, participants needed to recall the AI priming condition they were in at the end of the survey. We kept a reminder of which condition they were in throughout the survey on the top of every email. Those who answered the incorrect condition were excluded from the study. After qualification checks, we had 147 participants with mean age of 43.9 years (range: 25-77) of which 67.4% were male, 30% were female, and 2.6% preferred not to say. 47.37% identified as White, 13.16% as Black or African American, 57.89% as Asian, 2.6% as Native Hawaiian or Pacific Islander and 2.6% chose not to disclose.

We also recruited 10 participants from a large public university in the US to administer the survey and conducted interviews to qualitatively understand participant responses to the online survey. We recruited these participants through convenience sampling by sharing the study on mailing lists for undergraduate and graduate students studying CS at the university and randomly selecting 10 participants; as a result, our participants are familiar with AI and writing assistance tools to formulate educated responses. This represents a demographic shift, with mean participant age 20.9 years (range: 18-25) of which 40 % were male and 60% were female. 30% identified as White, 10% as Black or African American and 60% as Asian. This population shift might impact the findings of the interviews as people from different regions and social backgrounds may have different Computer Attitude and AI attitude scores. However, we found interview participants' survey responses to be directionally similar to those of MTurk recruited participants. This is supported by our regression result (Table 2) that shows gender, age and ethnicity are not significant independent variables for trustworthiness. This means that the qualitative findings can be applied to the general audience.

4.2 Experimental validation

Under the no-AI priming condition, we performed a validity check of our experimental design. The participants for the validity check were a different group of 24 Master-Mturkers from the main group of 147 participants. First, we asked "Do you think this email was (1) 'Definitely Human-written' to (6) 'Definitely AI-generated"' following [28] to calculate AI score. Second, we asked "*How emotionally involving would you rate this communication to be*" on a 1-6 Likert scale.

We found that the mean AI score was 2.42 with std. error .097, 95% CI [2.23-2.61] and a median of 2.00. This showed that participants generally thought of emails as human-written without receiving any priming prompt. This was expected because we wrote the emails ourselves. We confirmed that without manipulation the email themselves are thought to be human-written. For the Interpersonal Emphasis check, we found that for the Scenario Product Inquiry, the mean Interpersonal Emphasis rating was 2.89 with std. error 0.161, and 95% CI [2.57-3.20]. For the Scenario Party

Invitation, the mean Interpersonal Emphasis rating was 3.80 with std. error 0.139, and 95% CI [3.53-4.08]. For the Scenario Consolation of Pet Loss, the mean Interpersonal Emphasis rating was 4.61 with std. error 0.135, and 95% CI [4.35-4.88]. As expected, participants rated the Scenario Product Inquiry low on Interpersonal Emphasis,the Scenario Party Invitation higher, and the Scenario Consolation of Pet Loss highest. Thus, we found that 1. without priming, our hand written emails were indeed thought to be non-AI by participants and 2. our assumptions about Interpersonal Emphasis levels of the three scenarios were confirmed by participants.

4.3 Survey Results

We received 1764 data points from 147 participants (49 participants for each AI condition). We calculated an overview of our data. The average duration for taking the survey was 19.54 minutes; the minimum and maximum completion time were 12.75 minutes and 38 minutes, respectively. The average trust score, subject expertise score, propensity to trust, computer attitude score, and AI attitude score were 3.81, 3.63, 3.51, 3.61, and 3.41, respectively (Table 3). Since all average scores were greater than neutral scores (3), we saw a general bias towards higher scores for all variables. Nevertheless, there was no floor or ceiling effect.

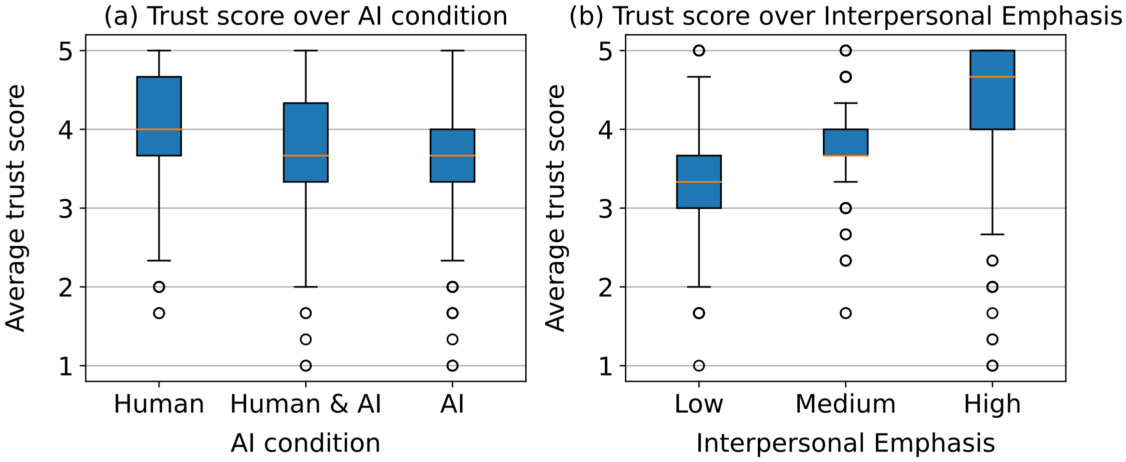

Fig. 1. Trustworthiness is calculated by an average of ratings across the three dimensions of trust. The blue boxes indicates the score region bounded by the 1st and 3rd Quartile. The horizontal orange lines are medians. The minimum and maximum trust scores were estimated by the equation 1st Quartile ±(3rd - 1st Quartile) * 1.5 .

We started by examining **RQ1: Does AI condition affect trustworthiness?** As we discussed earlier, previous research suggests that people tend to trust AI-generated content less than humans. We hypothesized that perceived trustworthiness would be lower under complete AI agency condition as compared to that under complete human agency condition (H1).

We plotted perceived trustworthiness score for each AI condition. Figure 1 (a) illustrates participants' trust ratings across AI conditions. From complete human agency to complete AI agency, the average trust scores decreased (4.02, 3.68, and 3.51), the median trust scores decreased (3.98, 3.66, and 3.63), and both the 1st and 3rd Quartile decreased. Figure 1 (a) confirms our H1, that the more AI-associated the condition is, the lower trustworthiness rating it receives.

To further understand this observation, during our interview we asked about participants' opinions on AI written messages. All 10 of them had reservations against the AI-generated content.

Participant 1 disliked the senders' lack of involvement in the writing process:

> "So for me, I am not too happy about the fact that the person used AI to write the email. I would expect them to be definitely more involved. I would be happier if things are more like raw and real."

Participant 10 thought the sender's AI usage takes away the authenticity of their condolences, therefore hindering their perceived benevolence,

> "I feel like seeing the following emails written by an advanced AI system, I feel like that kind of takes away from it. I feel like that makes it less authentic, like if someone's like, 'Oh, I'm sorry for your loss,' and you see... sent by a robot, it's like, okay."

Not only did participants dislike receiving an AI-generated message, 8 out of 10 of them rejected using AI tools to write their own emails.

Participant 3 remarked that they wish to follow a set of communication principles that the receiver expects (integrity),

> "Even if I like a smart suggestion, I just don't click it. It feels inauthentic, because I'm ... Okay. If I click this and send it to the person, they're going to think that this is me responding to them when ... And that I'm actually putting up an effort to hold a conversation with them. When in reality, I'm just clicking the button to say, 'Okay, check. I responded to this person. Now they're out of my notifications."'

We move on to **RQ2: Does Interpersonal Emphasis affect trustworthiness under different AI conditions?** As we discussed earlier, people tend to trust AI-generated content more when the nature of content is more fact driven and unrelated to interpersonal dynamics. Thus we hypothesized that under the complete AI agency condition, emails with higher Interpersonal Emphasis levels would show lower perceived trustworthiness score. Moreover, this effect would be reversed for emails in the complete human agency condition (H2).

We first calculated the average trust scores under different Interpersonal Emphasis Scenarios (Figure 1 (b)). From Low emphasis to High emphasis, the average trust scores increased (3,82, 4.01, and 4.45), the median trust scores increased, (3.56, 3.89, and 4.68), and both the 1st and 3rd Quartile increased. Figure 1 (b) shows that people trust an email more when its topic is more emotional. We further explore this effect with Figure 2.

Figure 2 combines Figure 1(a) and (b) to illustrate the relationship between an AI condition and Interpersonal Emphasis in regards to trust score. For all AI conditions, the higher the Interpersonal Emphasis, the higher the trust score. Note that even when people were told an email was written by an AI, they still trusted emails concerning a Consolation of Pet Loss more than those about Product Inquiries. This overthrows our H2. We suspect that people are more trusting of AI-generated content that appears more intimate.

To further understand this observation, we asked interview participants to explain their thought process behind the choice of ratings for the trustworthiness dimensions involving benevolence, ability, and integrity. We find that participants almost exclusively focused on the content of the email and the social context surrounding them while answering trustworthiness-scale questions with minimal or no regard towards the preceding AI priming prompt.

This includes the writing style or the tone of the message. For instance, participant 7 said,

> "I am basing it mostly on the tone of it. And how casual versus sincere they seemed."

Besides the writing style, participants valued the level of details, which improves the perceived benevolence of the sender. Participant 9 mentioned the specifics of the email as a way to understand that the sender cares,

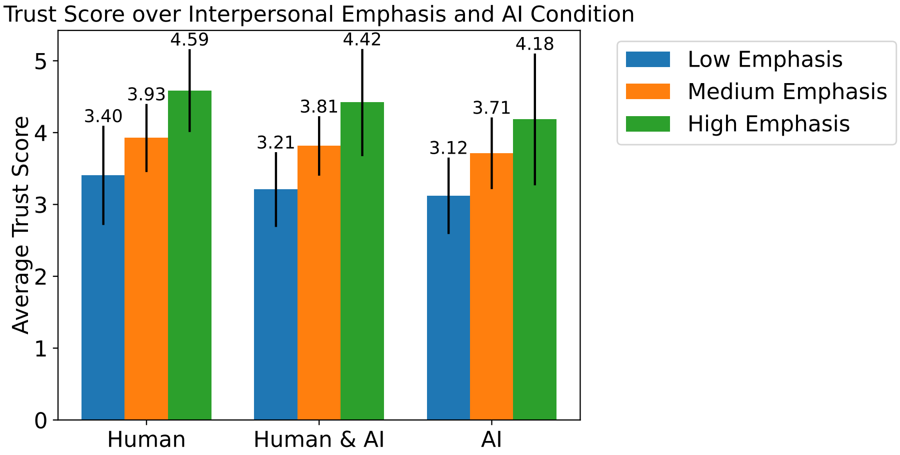

Fig. 2. Average trust scores for all Interpersonal Emphasis levels based on different AI conditions. Different Interpersonal Emphasis Scenarios are indicated by different colors. The float values above each bar represent the average trust score. The vertical black lines indicate the 95% confidence intervals.

> "I think the email sounds like they understand, because there is names, there is a location, there is also like... Some kind of life goals. And then usually they say something like call me or, I'm here for you."

Details about the logistics of the party also boost the perceived integrity of the sender. Participant 2 said,

> "I don't think she's hiding anything because she told him the details of where it is and when it is. So yeah, she told the person everything they need to know."

Overall, participants thought that the email should sound formal for Scenario product inquiries, and conversational for the Scenario Party Invitation and the Consolation of Pet Loss. Participants saw typos can be seen as either a good or bad sign of trustworthiness. Longer emails were preferred because they contain more details. Finally, social cues like the senders' financial status are used to evaluate their ability. Participant 4 remarked,

> "A fundraiser seemed like the company's flush with cash. They're going to celebrate it. Definitely capable of hosting."

However, at the same time, prevalent thoughts about the use of AI under high interpersonal emphasis conditions remain as we hypothesized. All 10 participants agreed that using an autonomous AI writing tool for highly interpersonal emails would be inappropriate. As participant 8 remarked,

> "If I were to receive condolences for any reason, and then later I were to find out that it wasn't really the person who wrote certain things... because I think I would take it to heart, whatever they said in the thing, so I wouldn't know. If I really took one sentence they wrote to heart and that was a sentence that wasn't even written by them or that was provided to them by the AI, I think that would affect me."

We observe a difference between how participants rated perceived trustworthiness by focusing primarily on the contents of the e-mail and ignoring the AI-priming condition while at the same time they deeply care about the use of AI in writing messages and specifically unanimously agree to not want to use them for high interpersonal emphasis messages such as consoling for the loss of a pet. We can reconcile this difference by the fact that participants found the quality of the e-mails to be so human-like that the messages didn't evoke the involvement of a machine or an AI when judging the message itself. Participants simply overlooked the AI priming prompt even though they read it. Instead, they focused on the social aspects based on the content of presented email messages.

Participant 5 perfectly captured this saying,

> "I just forgot I have the impression in my mind that those messages are written with the help of [an AI] system [...] while I was reading the messages because it felt quite natural. And they use this word like 'haha.' I thought that was quite funny. And the one talking about their past or, 'I remember I was visiting your place,' and then 'Your pet was on my lap'... I just felt like that just felt written by a real person. I couldn't think of other possibilities. Yeah, so I totally forgot that was the help of the system. It's quite amazing."

Finally we address **RQ3: Is trustworthiness impacted by subject expertise, AI attitude, computer attitude and/or general propensity to trust?** by performing a linear regression on our data to validate our previous findings and answer our RQ3. Table 2 demonstrates the relationship between the dependent variable (trust score) and independent variables. With trust as the dependent variable, we built three models for linear regression. Model 1 includes age, gender, and ethnicity as controls. This model serves as our baseline. Model 2 features two additional independent variables: Interpersonal Emphasis and AI condition. Lastly, we used all independent variables for Model 3.

In Model 1, we did not find age, gender, and ethnicity to be significantly correlated to perceived trustworthiness score. Model 2 shows that the Interpersonal Emphasis and AI condition are statistically significant with p values of 0.000. Interpersonal Emphasis has the most influence on trust score with a Beta of 0.450, while AI condition negatively affects trust score with a Beta of -0.225. This confirms the findings in Figure 1, showing that an Interpersonal Emphasis positively affects trust score and AI Condition negatively affects trust score.

Finally, in Model 3 We found subject expertise to be significantly correlated with perceived trustworthiness score. It is positively correlated with trust score with a Beta of 0.137. On further running a between-subjects analysis of variance, we find F(4,1759) = 88.927 $p < 0.001$. Thus we reject the null hypothesis and conclude that subject expertise influences the perceived trustworthiness score. Furthermore, we did not find AI attitude, computer attitude, and general propensity to trust to be significantly correlated to perceived trustworthiness score. Thus we fail to reject the null hypothesis. However, we did find directionally positive results with Beta = 0.016 (AI Attitude), 0.002 (computer attitude), 0.023 (propensity to trust).

Our interviews also revealed that subject expertise plays a role in influencing trustworthiness decisions. Eight out of ten participants mentioned, unprompted, how they used their past experiences to understand and judge the trustworthiness of email messages. Participant 7 said,

> "For some of the ones where they're asking about a product or trying to buy something, I've definitely done the same thing as in asking a bunch of questions, especially with that initial email and just giving them all that information. So for the snowflake email, I would do the same thing as saying, Hey, this is what we'd had in the past business. We want something comparable."

Participant 5 mentioned that his attachment to the contents of the message was influential in informing his trustworthiness ratings,

> "I love dogs. So I think that is pretty influential on my answers. And I do have a really close friend back in college, she lost her dog. One of her dogs, so I can emphasize a little bit, with those kinds of messages."

Model	**R^2**	**Variables**		**Standardized Coefficients**	**P Value**
			Std. Error	Beta	
1	0.017	(Constant)	0.220		0.000
		Age	0.005	-0.068	0.222
		Gender	0.086	-0.079	0.135
		Ethnicity	0.028	-0.077	0.148
2	0.271	(Constant)	0.202		0.000
		Age	0.004	-0.036	0.462
		Gender	0.075	-0.036	0.436
		Ethnicity	0.025	-0.080	0.082
		Interpersonal Emphasis	0.066	**0.367**	**0.000**
		Condition	0.067	**-0.308**	**0.000**
		Interperonal Emphasis * Condition	0.051	0.136	0.118
3	0.289	(Constant)	0.321		0.000
		Age	0.004	-0.038	0.452
		Gender	0.081	-0.050	0.317
		Ethnicity	0.026	-0.089	0.064
		Interpersonal Emphasis	0.066	**0.334**	**0.000**
		Condition	0.072	**-0.282**	**0.000**
		Interperonal Emphasis * Condition	0.050	0.137	0.111
		Subject Expertise	0.028	**0.137**	**0.003**
		Propensity to Trust	0.037	0.023	0.656
		Computer Attitude	0.071	-0.002	0.976
		AI Attitude	0.059	-0.016	0.802

Table 2. Coefficient table from three linear regression models where the dependent variable is average trust score.

In summary:

- **H1 Confirmed:** Messages under the complete AI-agency condition rate low on trustworthiness as compared to those with no priming and complete human agency.
- **H2 Disconfirmed:** Regardless of the AI condition, messages with higher Interpersonal Emphasis levels were perceived as more trustworthy.
- **H3 Disconfirmed:** We did not find AI attitude, computer attitude, and general propensity to trust to be significantly correlated to perceived trustworthiness score.
- **H4 Confirmed:** Subject expertise positively impacts perceived trustworthiness.

The confirmation of our hypothesis 1 supports previous studies that concluded perceived algorithmic reply use negatively affects the sender's trustworthiness [24, 28]. Our qualitative results provide explanations for this effect, showing that participants found the usage of AI tools (1) demonstrates the lack of effort and (2) takes away the emotional weight of condolences. Our finding of H2 opposes the belief that states AI-generated information with lower interpersonal emphasis is perceived to be more warranted [13] given that machines are seen as more objective than humans [54]. The interviewee's rationale on how they determine trustworthiness demonstrates the social information processing theory [57] that suggests individuals view language content and style characteristics as primary conduits of interpersonal information. Participants focus on paralinguistic cues [55, 56] instead of the AI condition. Finally, different from previous AIMC literature [24, 28], we quantitatively and qualitatively highlight the effect of subject expertise on perceived trustworthiness (H4). This validates the theory on intrinsic trust [27] that says such trust can only be gained when the user has background knowledge on what behavior is trustworthy for a given task the AI performs.

5 DISCUSSION

In this section, we discuss and explain our findings about participants' perceived trustworthiness of e-mail messages. We also highlight theoretical and design implications of this research, discuss limitations, and identify directions for future work.

We observed that there was a significant main effect of the AI-priming condition on perceived trustworthiness score. While the effect size was not large, it did reflect an apparent decrease in trust for messages perceived to be written by an AI system. These findings were corroborated by our interviews where participants were optimistically cautious about the use of AI systems. While qualitatively we found that participants focused on the content of the messages when rationalizing their decisions for trustworthiness ratings, it seemed there was a subtle impact of the AI priming condition that lowered overall scores with increase in perceived AI agency.

Our findings about the effect of AI priming coincide with Jakesch et al. [28]. They study the impact of AI mediation on trustworthiness by exposing participants to either solely AI priming or mixed AI & Human priming. They find that "participants were willing to accept and trust the AI-mediation, possibly due to the uniform application of the [AI priming] technology." Our study offers a potential broader understanding to their study that while participants might be generally highly trusting of an AI-primed source if all of them are purported to be by AI, they might still be subtly affected by its presence which would lead to overall lower perceived trustworthiness. These findings concur with [24] who find that "*as participants think that their partner is using more algorithmic responses, they perceive them as less cooperative and feel less affiliation towards them.*"

Besides the effect of AI priming, we also examined the impact of Interpersonal Emphasis Scenarios. In our quantitative finding, participants were more trusting of messages with higher Interpersonal Emphasis under all AI conditions (Figure 2). However, qualitatively no participants accept AI's writing in the Scenario Consolation of Pet Loss. There exists a disparity between what people say about AI-generated-emails and how they actually react to it.

This finding might be explained by the Hyperpersonal Model [55, 56] that suggests participants use para-linguistic cues to access their communication partners in CMC. This concurs with the findings from our interviews, where participants mainly rationalized their trustworthiness ratings based on the content of the message and its corresponding social context. Walther [55, 56] add, "*with fewer cues to base their perceptions, receivers have to 'fill in the gaps' of their understanding of the other interactant.*" We observed in our interviews that participants filled in these gaps by judging the sender based on the level of details, the tone, and other social cues. When participants focused on the content instead of the AI priming conditions, they trusted emails with higher Interpersonal Emphasis more because these emails contains rich para-linguistic or social cues such as a caring tone.

The Hyperpersonal Model may also explain why we observed a significant main effect of subject expertise on perceived trustworthiness score. Since the emails were written by humans with background knowledge of the emails' topics, participants with greater overall familiarity about these topics were more likely to trust it based on content cues available than those who did not. It also follows [6] who find that "*individuals might be motivated to examine relevant information as a strategy to minimize the implicit doubt that accompanies an inconsistency between explicit and implicit self-conceptions.*" This might explain why participants elicited details about the contents of the message when justifying their trustworthiness ratings while knowing the involvement of AI in writing the message.

5.1 Implication and Design Recommendations

In this study, we provide a foundation for understanding how people perceive AI writing tools, and more broadly, AI-mediated communications. Our study indicates that receivers of communication would remain oblivious to the fact that an AI system might have been used to augment the communication to a high degree. Such knowledge of the extent of AI mediation is crucial for users' underlying needs to achieve positive experiences [61].

Our work raises a serious ethical dilemma. As we design tools and services that allow the use of AI to mediate communication, should we enforce a preamble that notifies receivers about the use of AI to augment the said communication? If so, to what extent and why? As we find, if receivers would be oblivious to such disclaimers with their focus on the quality of the message, why should this policy be enforced? At the same time, the use of AI-mediated communication can allow senders to better express their emotions and show empathy towards the receiver which enhances the self-presentation offered by CMC.

However, as Participant 7 remarked, "*it freaks me out a little bit, AI being able to replicate human subjects or some more casual conversation that way.*" If it's an eerie phenomenon for the receiver, the sender should educate the receiver about the use of such an AI system. As we find through our study, if receivers are explained about the use of such AI systems with high degrees of autonomy, they disapprove it for messages with high interpersonal emphasis and would reflect poorly on the sender. Does this mean that the use of such AI systems should be avoided? Or should their disclosure be avoided? It is hard to establish a clear answer to this questions. As societal expectations for the use of such systems develop, it would perhaps become clear. Until then, it remains a moral gray area.

Our work also raises wider societal implications on the effectiveness and the use of AI mediated communication for protected categories such as lobbying for votes or for influencing purchasing behaviors of customers. If an AI mediated communication increases the efficacy of humans engaging in such activities [8, 49], our work suggests that it might be a logical decision to use such AI mediation as receivers would not perceive such mediation negatively. But it raises moral and ethical concerns about the use of semi automation for influencing citizens' behavior. These concerns are valid across a wide range of civic projects, including healthcare [34], immigration [11], autonomic vehicles [36], etc. For instance, California's SB 1001 "Bolstering Online Transparency" or the B.O.T law requires all bots that attempt to influence California residents' voting or purchasing behaviors to conspicuously declare themselves as bots [23]. However, AI-mediation lies in various degrees between computer mediation and full autonomy. It lies at the ambiguous line between a "bot" and "not-a-bot". Therefore, policy makers must decide on a rather nuanced scale what degrees of AI mediated communication might be allowed for different protected activities.

5.2 Limitations

This work is subject to several limitations. First, we asked participants to engage with senders they did not know. This meant that they ignored an important social cue about caring who the sender is. This can harm the study's ecological validity as participants' familiarity with the sender can be influential to the results in high interpersonal situations. The adoption of the participants' friends as the senders may positively increase trustworthiness. The increase may be larger for higher interpersonal scenarios. Future studies might investigate the impact of AI-MC on perceived trustworthiness against different levels of interpersonal involvement by participants.

Second, the wizard-of-oz setup limits the generalizability of the study's results. Our study does not mimic AI-Mediated Communication with the current state of the technology. Instead, it evaluates how AI would be perceived in a future time where the AI's response is indistinguishable from that of humans. A follow-up study may evaluate how different contemporary algorithms like GPT-3 actually generate the replies and their perceived trustworthiness.

Third, we chose emails as the communication medium. However, people tend to use instant messaging platforms to communicate with friends and for inviting someone to a party instead of using email messages. We do not know if social cues might be weighed differently when using AI-mediated communication via other communication mediums. Future studies may extend our work to other media such as instant messages (on different platforms), long form communication (via blog posts, newsletters, etc.), and multi-modal communication (such as TikTok, Instagram stories etc).

6 CONCLUSION

Our study investigates people's perceptions of AI-mediated communication in the context of writing emails. We introduced three Interpersonal Emphasis Scenarios to better categorize the messages. In each scenario, we explored how people's trust towards the messages changed over different AI conditions. We found that people's trust decreased as the perceived sender of the email shifted from human to AI. We also present the interesting phenomenon that people trust AI's writing in Scenario Consolation of Pet loss more than that in Scenario Product Inquiry. In the Mturk portion of our study, participants found more highly interpersonal messages to be more trustworthy. In our follow-on interviews, when explicitly asked about whether they would welcome an AI-written message on a highly personal topic, users reacted negatively. The same message evoked opposite reactions. We suspect that in the MTurk part of our study, participants didn't focus attention on who wrote the message as much as on how well it was written. We found that most participants used a combination of para-linguistic and social cues to determine trustworthiness.

The results of this study address the distinction between what people say about AI and how they actually react to it. This suggests that AI writing-assistance tools will be accepted over time if they work well. In this project, we chose to use all human-written text to better isolate people's feelings about who wrote a text rather than what the text said. In future work, we hope to study people's reactions to texts written by AIs. Both the technology of AI and people's perceptions of it are rapidly evolving. Those phenomena are inter-twined– how well the technology performs changes people's feelings about its use. This research serves as a snapshot at one moment in time, and it would be worthwhile to track how these perceptions evolve over time.

REFERENCES

[1] Ritu Agarwal and Jayesh Prasad. 1999. Are individual differences germane to the acceptance of new information technologies? *Decision sciences* 30, 2 (1999), 361–391.

[2] Mary D Salter Ainsworth, Mary C Blehar, Everett Waters, and Sally N Wall. 2015. *Patterns of attachment: A psychological study of the strange situation.* Psychology Press.

[3] Nancy K Baym. 1995. The emergence of community in computer-mediated communication. (1995).

[4] Joyce Berg, John Dickhaut, and Kevin McCabe. 1995. Trust, reciprocity, and social history. *Games and economic behavior* 10, 1 (1995), 122–142.

[5] John Bowlby. 1969. *Attachment and Loss: Attachment; John Bowlby.* Basic books.

[6] Pablo Briñol, Richard E Petty, and S Christian Wheeler. 2006. Discrepancies between explicit and implicit self-concepts: Consequences for information processing. *Journal of personality and social psychology* 91, 1 (2006), 154.

[7] Tom B Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *arXiv preprint arXiv:2005.14165* (2020).

[8] Daniel Buschek, Martin Zürn, and Malin Eiband. 2021. The Impact of Multiple Parallel Phrase Suggestions on Email Input and Composition Behaviour of Native and Non-Native English Writers. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 732, 13 pages. https://doi.org/10.1145/3411764.3445372

[9] Alex Calderwood, Vivian Qiu, Katy Ilonka Gero, and Lydia B Chilton. 2020. How Novelists Use Generative Language Models: An Exploratory User Study.. In *HAI-GEN+ user2agent@ IUI.*

[10] Karen S Cook, Toshio Yamagishi, Coye Cheshire, Robin Cooper, Masafumi Matsuda, and Rie Mashima. 2005. Trust building via risk taking: A cross-societal experiment. *Social psychology quarterly* 68, 2 (2005), 121–142.

[11] Eric Corbett and Christopher Le Dantec. 2021. Designing Civic Technology with Trust. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 173, 17 pages. https://doi.org/10.1145/3411764.3445341

[12] Nils Dahlbäck, Arne Jönsson, and Lars Ahrenberg. 1993. Wizard of Oz studies: why and how. In *Proceedings of the 1st international conference on Intelligent user interfaces*. 193–200.

[13] David C DeAndrea. 2014. Advancing warranting theory. *Communication Theory* 24, 2 (2014), 186–204.

[14] John December. 1996. Units of analysis for Internet communication. *Journal of Computer-Mediated Communication* 1, 4 (1996), JCMC143.

[15] Jens Edlund, Joakim Gustafson, Mattias Heldner, and Anna Hjalmarsson. 2008. Towards human-like spoken dialogue systems. *Speech communication* 50, 8-9 (2008), 630–645.

[16] Amanda J Ferguson and Randall S Peterson. 2015. Sinking slowly: Diversity in propensity to trust predicts downward trust spirals in small groups. *Journal of Applied Psychology* 100, 4 (2015), 1012.

[17] Jeffrey T Hancock, Mor Naaman, and Karen Levy. 2020. AI-Mediated Communication: Definition, Research Agenda, and Ethical Considerations. *Journal of Computer-Mediated Communication* 25, 1 (2020), 89–100.

[18] Jeffrey T. Hancock, Jennifer Thom-Santelli, and Thompson Ritchie. 2004. *Deception and Design: The Impact of Communication Technology on Lying Behavior*. Association for Computing Machinery, New York, NY, USA, 129–134. https://doi.org/10.1145/985692.985709

[19] Jeffrey T. Hancock, Catalina Toma, and Nicole Ellison. 2007. *The Truth about Lying in Online Dating Profiles*. Association for Computing Machinery, New York, NY, USA, 449–452. https://doi.org/10.1145/1240624.1240697

[20] Russell Hardin. 2002. *Trust and trustworthiness*. Russell Sage Foundation.

[21] Anthony Hartley and Donia Scott. 2001. Evaluating text quality: judging output texts without a clear source. In *Proceedings of the ACL 2001 Eighth European Workshop on Natural Language Generation (EWNLG)*.

[22] Susan C Herring. 1996. *Computer-mediated communication: Linguistic, social, and cross-cultural perspectives*. Vol. 39. John Benjamins Publishing.

[23] Matthew Hines. 2019. I smell a bot: California's SB 1001, free speech, and the future of bot regulation. *Hous. L. Rev.* 57 (2019), 405.

[24] Jess Hohenstein, Dominic DiFranzo, Rene F Kizilcec, Zhila Aghajari, Hannah Mieczkowski, Karen Levy, Mor Naaman, Jeff Hancock, and Malte Jung. 2021. Artificial intelligence in communication impacts language and social relationships. *arXiv preprint arXiv:2102.05756* (2021).

[25] Jess Hohenstein and Malte Jung. 2018. AI-Supported Messaging: An Investigation of Human-Human Text Conversation with AI Support *(CHI EA '18)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3170427.3188487

[26] Jess Hohenstein and Malte Jung. 2020. AI as a moral crumple zone: The effects of AI-mediated communication on attribution and trust. *Computers in Human Behavior* 106 (2020), 106190.

[27] Alon Jacovi, Ana Marasovic, Tim Miller, and Yoav Goldberg. 2020. Formalizing Trust in Artificial Intelligence: Prerequisites, Causes and Goals of Human Trust in AI. *CoRR* abs/2010.07487 (2020). arXiv:2010.07487 https://arxiv.org/abs/2010.07487

[28] Maurice Jakesch, Megan French, Xiao Ma, Jeffrey T Hancock, and Mor Naaman. 2019. AI-mediated communication: How the perception that profile text was written by AI affects trustworthiness. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.

[29] Sara Kiesler, Jane Siegel, and Timothy W McGuire. 1984. Social psychological aspects of computer-mediated communication. *American psychologist* 39, 10 (1984), 1123.

[30] Toko Kiyonari, Toshio Yamagishi, Karen S Cook, and Coye Cheshire. 2006. Does trust beget trustworthiness? Trust and trustworthiness in two games and two cultures: A research note. *Social psychology quarterly* 69, 3 (2006), 270–283.

[31] Bran Knowles, Mark Rouncefield, Mike Harding, Nigel Davies, Lynne Blair, James Hannon, John Walden, and Ding Wang. 2015. Models and Patterns of Trust. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work amp; Social Computing* (Vancouver, BC, Canada) *(CSCW '15)*. Association for Computing Machinery, New York, NY, USA, 328–338. https://doi.org/10.1145/2675133.2675154

[32] Cliff A.C. Lampe, Nicole Ellison, and Charles Steinfield. 2007. *A Familiar Face(Book): Profile Elements as Signals in an Online Social Network*. Association for Computing Machinery, New York, NY, USA, 435–444. https://doi.org/10.1145/1240624.1240695

[33] Laura Larrimore, Li Jiang, Jeff Larrimore, David Markowitz, and Scott Gorski. 2011. Peer to peer lending: The relationship between language features, trustworthiness, and persuasion success. *Journal of Applied Communication Research* 39, 1 (2011), 19–37.

[34] Min Kyung Lee and Katherine Rich. 2021. Who Is Included in Human Perceptions of AI?: Trust and Perceived Fairness around Healthcare AI and Cultural Mistrust. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 138, 14 pages. https://doi.org/10.1145/3411764.3445570

[35] Gale M Lucas, Jonathan Gratch, Aisha King, and Louis-Philippe Morency. 2014. It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior* 37 (2014), 94–100.

[36] Stefanie M. Faas, Johannes Kraus, Alexander Schoenhals, and Martin Baumann. 2021. Calibrating Pedestrians' Trust in Automated Vehicles: Does an Intent Display in an External HMI Support Trust Calibration and Safe Crossing Behavior?. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 157, 17 pages. https://doi.org/10.1145/3411764.3445738

[37] Xiao Ma, Jeffery T. Hancock, Kenneth Lim Mingjie, and Mor Naaman. 2017. Self-Disclosure and Perceived Trustworthiness of Airbnb Host Profiles *(CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 2397–2409. https://doi.org/10.1145/2998181.2998269

[38] François Mairesse and Marilyn Walker. 2006. Automatic recognition of personality in conversation. In *Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers*. 85–88.

[39] Franc Mairesse, Marilyn Walker, et al. 2006. Words mark the nerds: Computational models of personality recognition through language. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 28.

[40] François Mairesse, Marilyn A Walker, Matthias R Mehl, and Roger K Moore. 2007. Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of artificial intelligence research* 30 (2007), 457–500.

[41] Roger C Mayer, James H Davis, and F David Schoorman. 1995. An integrative model of organizational trust. *Academy of management review* 20, 3 (1995), 709–734.

[42] D Harrison McKnight, Vivek Choudhury, and Charles Kacmar. 2002. Developing and validating trust measures for e-commerce: An integrative typology. *Information systems research* 13, 3 (2002), 334–359.

[43] Gary S Nickell and John N Pinto. 1986. The computer attitude scale. *Computers in human behavior* 2, 4 (1986), 301–306.

[44] James W Pennebaker, Matthias R Mehl, and Kate G Niederhoffer. 2003. Psychological aspects of natural language use: Our words, our selves. *Annual review of psychology* 54, 1 (2003), 547–577.

[45] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research* 21, 140 (2020), 1–67.

[46] Ronald E Robertson, Alexandra Olteanu, Fernando Diaz, Milad Shokouhi, and Peter Bailey. 2021. "I Can't Reply with That": Characterizing Problematic Email Reply Suggestions. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 724, 18 pages. https://doi.org/10.1145/3411764.3445557

[47] Ronald E Robertson, Alexandra Olteanu, Fernando Diaz, Milad Shokouhi, and Peter Bailey. 2021. "I Can't Reply with That": Characterizing Problematic Email Reply Suggestions. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–18.

[48] Julian B Rotter. 1971. Generalized expectancies for interpersonal trust. *American psychologist* 26, 5 (1971), 443.

[49] Quentin Roy, Sébastien Berlioux, Géry Casiez, and Daniel Vogel. 2021. Typing Efficiency and Suggestion Accuracy Influence the Benefits and Adoption of Word Suggestions. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 714, 13 pages. https://doi.org/10.1145/3411764.3445725

[50] Astrid Schepman and Paul Rodway. 2020. Initial validation of the general attitudes towards Artificial Intelligence Scale. *Computers in Human Behavior Reports* 1 (2020), 100014.

[51] Weiyan Shi, Xuewei Wang, Yoo Jung Oh, Jingwen Zhang, Saurav Sahay, and Zhou Yu. 2020. Effects of Persuasive Dialogues: Testing Bot Identities and Inquiry Strategies. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Apr 2020). https://doi.org/10.1145/3313831.3376843

[52] Peter Stone, Rodney Brooks, Erik Brynjolfsson, Ryan Calo, Oren Etzioni, Greg Hager, Julia Hirschberg, Shivaram Kalyanakrishnan, Ece Kamar, Sarit Kraus, et al. 2016. Artificial intelligence and life in 2030: the one hundred year study on artificial intelligence. (2016).

[53] Sandeep Subramanian, Raymond Li, Jonathan Pilault, and Christopher Pal. 2019. On extractive and abstractive neural document summarization with transformer language models. *arXiv preprint arXiv:1909.03186* (2019).

[54] S. Shyam Sundar. 2007. The MAIN Model : A Heuristic Approach to Understanding Technology Effects on Credibility.

[55] Joseph B Walther. 1996. Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction. *Communication research* 23, 1 (1996), 3–43.

[56] Joseph B Walther. 2011. Theories of computer-mediated communication and interpersonal relations. *The handbook of interpersonal communication* 4 (2011), 443–479.

[57] Joseph B. Walther. 2015. *Social Information Processing Theory (CMC)*. John Wiley Sons, Ltd, 1–13. https://doi.org/10.1002/9781118540190.wbeic192 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118540190.wbeic192

[58] Jennifer Wang and Angela Moulden. 2021. *AI Trust Score: A User-Centered Approach to Building, Designing, and Measuring the Success of Intelligent Workplace Features*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3411763.3443452

[59] Darcy Warkentin, Michael Woodworth, Jeffrey T. Hancock, and Nicole Cormier. 2010. Warrants and Deception in Computer Mediated Communication. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work* (Savannah, Georgia, USA) *(CSCW '10)*. Association for Computing Machinery, New York, NY, USA, 9–12. https://doi.org/10.1145/1718918.1718922

[60] Rui Yan. 2018. " Chitty-Chitty-Chat Bot": Deep Learning for Conversational AI.. In *IJCAI*, Vol. 18. 5520–5526.

[61] Xi Yang and Marco Aurisicchio. 2021. Designing Conversational Agents: A Self-Determination Theory Approach. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 256, 16 pages. https://doi.org/10.1145/3411764.3445445

A VARIABLE STATISTICS

In Table 3, we have N = 288 for the validation study and N = 1764 for the large scale study. Average Trust Score, Subject Expertise, Propensity to Trust, Computer Attitude and AI attitude are on five-point Likert scale. Validation questions for AI Condition and Interpersonal Emphasis have a range from one to six. Note that the mean for experimental validation and main experiment are not comparable because participants in the former study were asked validation questions is incompatible with our wizard-of-oz approach.

Variable Statistics	Experimental Validation		Main Experiment		Minimum	Maximum
	N	Mean	N	Mean		
Average Trust Score	288	3.86	1764	3.81	1	5
Subject Expertise	288	3.48	1764	3.63	1	5
Propensity to Trust	288	3.33	1764	3.51	1	5
Computer Attitude	288	3.25	1764	3.61	1	5
AI Attitude	288	3.19	1764	3.41	1	5
AI Condition Validation	288	2.42			1	6
Interpersonal Emphasis Validation	288	3.77			1	6

Table 3. Mean and range of variables for the two quantitative studies.

B INTERACTION EFFECT BETWEEN AI CONDITION AND INTERPERSONAL EMPHASIS

We run a between subject univariate analysis of variance to understand the interaction effect between AI priming and interpersonal emphasis condition. We observe a significant interaction effect of interpersonal emphasis condition and AI priming on perceived trustworthiness score with F(4,1759)= 2.986, p=0.019. Figure 3 shows the interaction.

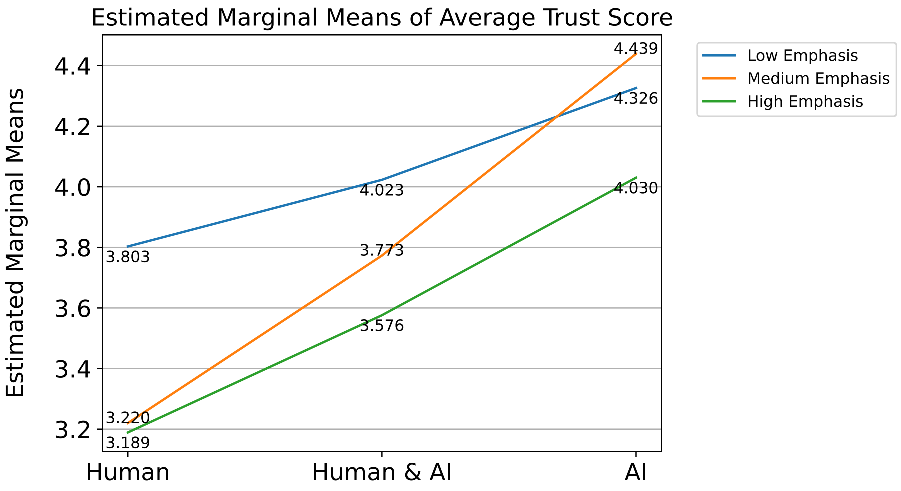

Fig. 3. Interaction graph of AI Condition and Interpersonal Emphasis