"Hi Alex" or "Dear Dr. Morgan"?: Exploring How AI-suggested Politeness Strategies Influence Email Writing and Social Perception Among Native and Non-native Speakers

ZIBO SELENA ZHANG, University of Waterloo, Canada
YUNA CHOI, University of Waterloo, Canada
ZIANG XIAO, Johns Hopkins University, USA
Q. VERA LIAO, Microsoft Research, Canada and University of Michigan, Ann Arbor, USA
JIAN ZHAO, University of Waterloo, Canada

As AI writing assistants are increasingly used for interpersonal communication, they may have profound impacts on interpersonal relationships. Politeness is one important aspect of social communication that is grounded in people's perceptions of relational dynamics and significantly shapes social interactions. We investigate how politeness strategies in AI-generated suggestions affect people's email writing and alter their perception of the social situation. Through a within-subject online study (N = 52), we found that human writers tend to mirror the type of politeness strategies used in the AI suggestions when writing their own messages. Non-native English speakers are more affected by AI compared to native speakers, and this greater susceptibility to AI influence is partly mediated by higher reliance on AI tools. In addition, writers' social perception is also influenced by AI. When writers are exposed to more deferential politeness strategy suggestions, they tend to perceive the social relationship as more distant. These findings highlight the need for better design of AI writing assistants that account for social contexts and individual differences.

CCS Concepts: • **Human-centered computing** → **Empirical studies in HCI**; **Empirical studies in collaborative and social computing**; • **Social and professional topics** → *Cultural characteristics.*

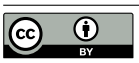
Additional Key Words and Phrases: computer-mediated communication, non-native speakers, large language models, writing assistant

ACM Reference Format:
Zibo Selena Zhang, Yuna Choi, Ziang Xiao, Q. Vera Liao, and Jian Zhao. 2026. "Hi Alex" or "Dear Dr. Morgan"?: Exploring How AI-suggested Politeness Strategies Influence Email Writing and Social Perception Among Native and Non-native Speakers. *Proc. ACM Hum.-Comput. Interact.* 10, 6, Article CSCW094 (October 2026), 33 pages. https://doi.org/10.1145/3816942

1 Introduction

In social interactions, people intentionally adopt communication strategies and language styles, including word choices and tones, to convey social intentions and shape social relationships [44]. Especially in initiating requests, expression of politeness is crucial and needs to be strategic for building positive relationships with the recipient and increasing the likelihood of a successful outcome [16]. Brown & Levinson's theory [11] outlines two types of politeness strategies in social

Authors' Contact Information: Zibo Selena Zhang, s95zhang@uwaterloo.ca, University of Waterloo, Waterloo, Ontario, Canada; Yuna Choi, yh6choi@uwaterloo.ca, University of Waterloo, Waterloo, Ontario, Canada; Ziang Xiao, ziang.xiao@jhu.edu, Johns Hopkins University, Baltimore, Maryland, USA; Q. Vera Liao, veraliao@umich.edu, Microsoft Research, Montreal, Quebec, Canada and University of Michigan, Ann Arbor, Ann Arbor, Michigan, USA; Jian Zhao, jianzhao@uwaterloo.ca, University of Waterloo, Waterloo, Ontario, Canada.



ACM 2573-0142/2026/10-ARTCSCW094
https://doi.org/10.1145/3816942

communication. *Negative politeness* aims to minimize the message receiver's feeling of being imposed upon, by incorporating indirect language or showing appreciation. *Positive politeness* aims to make the message receiver feel good about themselves, by using friendly language, compliments, or small talk. The use of these strategies is based on the communicator's evaluation of the social situation to achieve a desirable outcome.

With the surge of generative AI, smart writing assistants have been increasingly adopted in social communications, including in the professional context. For example, AI writing assistants can help generate or refine emails, LinkedIn messages, or instant messages on platforms like Slack. Previous research studying this kind of AI-mediated communication (AI-MC) has warned that AI tools could influence writers' language choices [5, 38], self-representation [46], and perceived agency [27], affect how receivers perceive the writers [25], and have lasting impacts such as shifting writers' opinions [24]. Building on this line of work, we critically examine how the use of AI writing assistants affects people's use of communication strategies, specifically politeness strategies. According to Brown & Levinson's theory [11], people choose politeness strategies according to their perception of social distance, power, and level of imposition (D, P, and R[1]). This suggests that the use of politeness strategies is guided by social perceptions. Thus, we additionally hypothesize that when AI shapes the use of politeness strategies in writing, people's social perceptions can be altered in the process. Our study on this topic reveals the risk of AI writing assistants distorting people's perception of others and themselves, as speculated conceptually in prior work [15, 18], and creating lasting social and psychological harms.

We also hypothesize that people with different language backgrounds might be impacted by AI in different ways. In this study, we focus on comparing native English speakers (NESs) and non-native English speakers (NNESs). How AI writing assistants might affect NNESs differently has been a frequently asked question. On the one hand, many believe that AI tools could have an equalizing effect by helping NNESs improve grammar and vocabulary [15], and it is possible that they could also help NNESs overcome the challenges of expressing intentions accurately [35], and conforming to social conventions [13]. On the other hand, a recent study shows that NNESs are more at risk of over-reliance on AI writing assistants [30]. It is therefore possible that NNESs will be more susceptible to the risks of being swayed by politeness strategies included in AI generation, and their social perception being distorted to a larger extent by writing with AI.

In summary, we ask the following research questions:

- **RQ1:** Do different politeness strategies included in AI writing assistant suggestions lead to different patterns of politeness strategies in writers' emails?
- **RQ2:** Do different politeness strategies included in AI writing assistant suggestions shift writer's perception of the social situation (referred to as social perception hereafter)?
- **RQ3:** Are non-native English speakers more subject to the influence of AI politeness strategies?

We study these questions in the context of AI-mediated email communication with professional requests. We conducted a within-subject online study ($N = 52$) to compare the effects of AI writing assistants that were set to favour different politeness strategies (positive strategies vs. negative strategies), and examine how NESs and NNESs might be impacted differently. The type of politeness strategies used by AI was set as the manipulated within-subject variable, while language background was a measured between-subject variable.

Our results show that politeness strategies used in emails are affected by AI assistants — participants used more positive politeness strategies when writing with AI using positive strategies (*positive-AI*) and more negative strategies with AI using negative strategies *(negative-AI*). This not only applies to the full email written with AI assistance, but also to the human-written components

[1]The use of the abbreviation 'D, P, R' is a direct adaptation from Brown & Levinson's original book

in the emails after removing AI suggestions. We also found evidence of AI politeness strategies shifting writers' social perception even after a one-off use, with *negative-AI* leading to a greater increase in perception of social distance, but no evidence suggests differential changes across AI politeness conditions in the other two dimensions (power and imposition). Because each participant experienced all AI conditions that differed only in politeness strategies, the observed differences can be attributed to the manipulation of AI politeness rather than confounds. In addition, we found significant differences between NESs and NNESs in being influenced by politeness strategies. That is, NNESs incorporated more positive politeness strategies in their emails compared to NESs, but we did not find a significant difference in the use of negative strategies. This differential effect on people with different language backgrounds is mediated by reliance, suggesting that NNESs are more impacted by AI because they rely on AI tools more than NESs. As language background is measured, group differences reflect correlations rather than causal effects. Based on these results, we discuss design implications for AI writing assistants supporting social communication, by calling out the need to be better aware of and align with the social contexts of communication, and to mitigate overreliance, especially for NNESs.

2 Background

2.1 AI-Mediated Writing for Communication

Writing is a major channel of social communication. Previous research has shown that a writer's word choices and language style are shaped by their personal characteristics, such as personality [29, 37, 43], and can also impact how the writer is perceived, influencing attributes like competence, dominance, and warmth [8]. The involvement of AI in the writing process means that the language used in social communication is no longer entirely produced by the writer. As a result, the connection between language style, writer characteristics, and receiver perception can be disrupted, calling for further investigation into how communication and social perceptions are reshaped with the widespread adoption of AI writing assistants.

Work in AI-mediated communication (AI-MC) has focused on these questions. This field is defined as "interpersonal communication in which an intelligent agent operates on behalf of a communicator by modifying, augmenting, or generating messages to accomplish communication goals" [18]. AI writing assistants impact both parties in the communication process: the sender, by influencing what they write and how they write, and the receiver, by shaping their perception of the writer based on the language and style used in the messages.

AI writing assistants can influence writers from multiple aspects of the writing process. Past studies revealed a positivity bias in AI-assisted writing, such as text messaging and restaurant reviews [5, 38]. This bias can be attributed to the fact that AI-generated texts tend to include a significantly higher proportion of positive language compared to human-written texts. AI suggestions are also found to shift topic selection in self-introductions [46], and change expressed opinions when writing argumentative essays with opinionated AI assistants [24]. In addition, the language style of AI — whether it uses confident, self-assured language or more hesitant language — influences people's feeling of control and ownership over the text they created [27]. Overall, previous studies have shown that different language styles used by AI can impact both how people construct messages and their perceptions.

AI writing assistants also influence how writers are perceived socially. AI can bring positive effects to communication, as communication partners who use AI assistants perceive the social relationship to be closer and more cooperative [21]. However, this work also indicated that when readers suspect that responses are formulated using AI, they are evaluated negatively. This is corroborated by another study suggesting that when Airbnb host profiles are suspected to be

written partly by AI, people's trust in the host is reduced [25]. There is mixed evidence suggesting that AI-generated content can also be perceived as less socially attractive [38].

Informed by this line of work, we are motivated to examine a complementary aspect: whether and how AI writing assistants' politeness strategies, as a form of communication strategy, influence writing content, as well as the writers' perception of the social situation.

2.2 Politeness Theory

While previous work in AI-MC examined readers' perceptions and writers' experiences, there is a gap in the literature exploring how AI-MC impacts communication strategies and relational perception. In this study, we focus on the communication strategies of expressing politeness. Politeness has been studied in the context of computer-mediated communication (CMC) [41]: negotiation of politeness through digital platforms like emails is different from face-to-face communication because of asynchronicity and anonymity [17]. Politeness has also been studied with AI, with work focused on using AI to detect politeness in text [57], or leveraging AI as a paraphrasing tool to facilitate accurate delivery of politeness [14].

AI's delivery of politeness strategies is often inaccurate, even with the recently developed generative language models. Qualitative insights from previous studies indicate that AI suggestions can be socially problematic, sometimes lacking appropriate politeness strategies (curt refusal) or using strategies that are inappropriate for the social scenario (giving overly informal suggestions for professional relationships) [49]. It is also suggested that the AI suggestions can be on relatively extreme and exaggerated ends, either too formal or too informal, and generally, AI suggestions are more suited for formal scenarios [15]. We note that, under Brown & Levinson's politeness theory (more below), *negative politeness strategies* tend to include more formal words and *positive strategies* include informal language. This suggests that current AI models tend to include politeness strategies that are more on the extreme ends, with a default tendency to suggest *negative politeness strategies*.

Our work is based on Brown & Levinson's politeness theory [11] drawing a link between social perception and the use of politeness strategies in social communication. Based on the theory, politeness stems from people's need to preserve their "face", which refers to people's need to have their actions unperturbed by others, and their need to maintain a positive, consistent self-image. However, in social interactions, it is inevitable to encounter situations where the face might be threatened, for example, making a request might involve imposing on other people. Politeness is used as a type of redressive action to minimize the face-threatening act (FTA) and to promote smooth social interactions.

Politeness strategies can be categorized into two types. *Negative politeness* is close to what we consider 'traditionally polite' and the goal is to minimize imposing on other people. This strategy is marked by some language characteristics, such as using indirect language (replacing 'will' with 'would', 'can' with 'could'), using words to minimize the imposition ('I need just *a little* of your time'), or acknowledging the impingement and apologizing ('I'm sorry to bother you, but...'). *Positive politeness* is used to maintain positive self-images and uses friendliness to avoid potential offence. This includes language features like compliments and exaggeration ('How absolutely marvellous'), including both parties in the activity ('Let's go to dinner'), or engaging in small talk ('How was your weekend?'). The full list of politeness strategies used in this paper is included in Appendix A.

As previous AI-MC work found that AI's language bias (e.g. positivity bias) can bias AI-MC content, we hypothesize that the politeness strategies of AI can have a similar effect:

H1: The politeness strategies used by AI writing assistants (*negative politeness* vs. *positive politeness*) increase the number of corresponding politeness strategies people include in the emails.

Brown & Levinson's theory postulates that the selection of politeness strategies is based on the communicator's estimation of the risk of face loss. Specifically, *positive politeness* is used when the writer perceives less risk, and *negative strategies* are used with more risk. The risk of face loss is estimated based on three factors, evaluated holistically. Distance (D) refers to how close or familiar two people feel toward each other, often based on how often they interact or exchange emotional support. Power (P) refers to how much one person can control or influence another. This can involve control over resources or the ability to shape others' beliefs and actions. Imposition (R) refers to how demanding the request is, and can be determined based on either the services (like time or effort) or goods requested. A higher risk of face loss is associated with higher evaluations of D, P, and R, while a lower risk of face loss corresponds to lower evaluations of these factors. The link between D, P, R perceptions and use of politeness strategies is supported by additional work, and shown to be applicable to professional settings [33].

When writers independently formulate their communication messages, evaluation of the situation impacts the choice of strategies. However, as AI writing assistants directly modify the written messages, we hypothesize that there could also be a reverse effect on social perception. The hypothesis is formalized below:

H2: AI writing assistants using positive versus negative politeness strategies can have differential impacts on people's social perception with regard to distance, power and imposition.

2.3 Non-native English Speakers and AI-assisted Social Communication

Compared to native English speakers (NESs), non-native English speakers (NNESs) face more challenges in professional communication, especially in email writing. The first contributing factor is the nature of email as a communication medium. This medium could be relatively unfamiliar or less favoured by NNESs from some cultures [56]. For example, people from East Asian backgrounds are familiar with high-context cultures, which rely more on implicit communication and contextual cues. As a result, they may find email communication as a low-context medium less intuitive. In addition, the conventions of email communication can be less well-defined, adding another layer of challenge for NNESs [9].

The second factor contributing to NNESs' communication challenges is pragmatic misunderstanding. Pragmatics is the study of how language and context are connected [34], and NNESs face more challenges in pragmatics than NESs. For example, NNESs are more likely to make pragmatic mistakes, like saying "please answer me as soon as possible" in emails to professors [13], which can be regarded as inappropriate in situations that require more deferential expressions. Other work shows that in incongruent social situations (when the social act and the status are not aligned), NNESs are less successful. The lack of success is not due to linguistic incompetence but because of the lack of pragmatic competence [6]. This pragmatic misunderstanding is connected to politeness theory, as people fail to comply with conventions because of their inaccurate evaluations of the situations (e.g., inaccurate perception of distance, power and imposition) [54].

AI writing assistants are expected to have an equalizing effect to help NNESs find the right words, improve grammar and vocabulary, and support cross-cultural communication [15]. However, recent studies find that, when using AI writing assistants, NNESs face challenges accessing context-appropriate suggestions [30], and they tend to over-rely on AI due to lower language proficiency and pragmatic understanding [30, 39]. This can lead to negative consequences, for example, the social intentions of NNESs are conveyed less accurately with machine translation, compared to those directly written in English [35]. Also, cultural diversity can be at risk with AI writing assistants, as NNESs adapt to foreign norms when writing with AI tools, both in the content and the language style [1].

Based on this previous work, we hypothesize that NNESs are more reliant on AI writing assistants and are more influenced by AI suggestions:

H3: Compared to NESs, NNESs are more susceptible to the influence of AI writing assistants on the use of politeness strategies in emails and their social perception.

H3a: NNESs are more influenced by AI in the use of politeness strategies in emails.

H3b: NNESs are more influenced by AI in social perception.

That is, we will test **H3** by analyzing the interaction effects of language background (*NES* versus *NNES*) and AI politeness condition (*positive-AI* versus *negative-AI*) on the dependent variables associated with **H1** and **H2**.

3 Method

To investigate our research questions (**RQ1–3**) and hypotheses (**H1–3**), we conducted an online study ($N = 52$) to examine the impact of AI POLITENESS on human email writing, social perception, and writing quality, and analyzed whether participants with different LANGUAGE backgrounds are impacted differently. We describe our experimental design in the following.

3.1 Study Design

We employed a within-subject design, with AI POLITENESS as the manipulated within-subject variable and LANGUAGE as the between-subject moderator variable. Each participant in the study completed three writing tasks under three conditions: one *baseline* condition in which participants wrote without AI, and two experimental conditions, *positive-AI* and *negative-AI*, in which participants wrote with AI tools that suggest either positive or negative politeness strategies. In each of the writing tasks, participants were asked to write based on one of three social scenarios, which are explained in more detail below in Section 3.2.3. We recruited both participants who are *native English speakers* (NESs) and *non-native English speakers* (NNESs). A methodological note on the study design is included in Appendix E.

3.2 Experimental Setup and Apparatus

3.2.1 Writing Task and Tool. Each participant completed three writing tasks, where they wrote emails based on three scenarios involving professional requests. They were asked to imagine themselves in the scenario as the person making the request, or to recall a previous experience in a similar situation.

In all three writing tasks, participants used three different variations of a writing tool that shared a similar interface, as shown in Figure 1. In the *baseline* condition, participants were provided with only a text box without AI suggestion features. They were discouraged from using any external AI tools, and were informed that if the use of such tools was detected, they may be denied remuneration. In the *positive-AI* and *negative-AI* conditions, participants were provided with two AI writing assistants that either favour positive politeness strategies or negative politeness strategies in the suggestions. The tools provided text continuation suggestions (approximately 20–30 words) based on existing content written by the user. The AI writing assistant was also prompted with the respective social scenario to ensure that the language suggested by the AI writing assistant was appropriately adapted to the scenario. We prompted the LLM with the scenarios to ensure that participants would receive useful suggestions for their tasks. In practice, people often get writing assistance from LLMs that are adapted to their task—either using domain-specific writing assistance for which the developers performed the adaptation, or providing the context in their own prompts. To avoid the individual variance in their prompting skills, we added the scenarios in the system prompt to create task-adapted writing assistance.

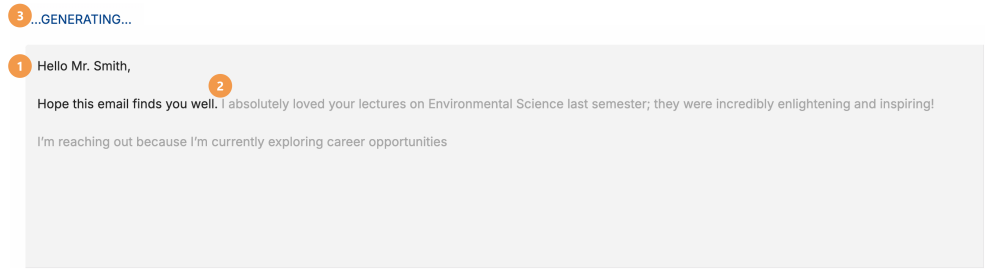


Fig. 1. The interface participants used in the writing tasks. 1: Participants start by writing the initial opening of the email; 2: AI suggestions can be called to continue the human-written components; 3: the "...Generating..." sign indicates that the system is working on generating suggestions.

The interface modelled AI assistant interactions in popular email applications and previous studies [27]. Specifically, text continuation suggestions were automatically generated after participants pressed the [space] or [enter] and paused typing for 1.5 seconds. This design ensured that suggestions do not disturb participants' writing flow when they are continuously typing. If participants were not satisfied with the suggestions, they could continue writing to ignore suggestions or press the [tab] key to re-generate suggestions. Participants were encouraged to use as many AI suggestions as they deemed appropriate.

Participants were motivated to produce high-quality emails through both social and monetary incentives. Before completing the baseline and AI writing tasks, they were provided with the following instruction: "Your email will be reviewed by individuals who could realistically be the recipients. These reviewers will rate the email's social appropriateness and the likelihood of providing a favorable response. The top 10% of participants with the highest ratings across three emails will receive an additional £3 reward." The inclusion of bonus rewards follows common practice in HCI to ensure quality control in studies that use online crowdsourcing platforms [19] [40] [10]. The performance-based bonus served as a monetary incentive, while emphasizing evaluation by plausible recipients functioned as a social incentive.

3.2.2 AI Model Configuration. The AI writing assistants in the *positive-AI* and *negative-AI* conditions differ in the politeness strategies included in the suggestions. These two AI assistants were configured differently to provide suggestions that either predominantly include positive politeness strategies or negative politeness strategies.

The suggestions were generated by calling the gpt-4o-2024-08-06 API. The difference between conditions was achieved through prompting. The prompts included the definition and examples of the politeness strategies based on [11] and [41]. The full list of politeness strategies with descriptions and examples is included in Appendix A.

To validate the politeness manipulation, we conducted two manipulation checks. In the first step, we examined whether emails generated using positive politeness prompts, negative politeness prompts, and the default GPT-4o model (without additional prompt modifications) were distinguishable on key manipulated variables at an overall level. The emails used in this test were generated in varied scenarios, including low-stakes interactions between friends, high-stakes email requests related to fundraising, and middle scenarios between the two extremes. The manipulation check was done through human rating on 12 AI-generated emails: 4 using the default GPT-4o model without additional prompts, 4 with positive politeness prompts, and 4 with negative politeness prompts. Two raters were provided with definitions and descriptions of positive and negative politeness

strategies and were instructed to categorize the emails into the three categories: default, positive, and negative. Rater 1 correctly categorized the emails with (75%) accuracy, while Rater 2 achieved (67%) accuracy. The most common error was misclassifying default emails as negative politeness emails. This is consistent with previous work [15] suggesting that GPT's default language style aligns more closely with negative politeness strategies. Both raters achieved (87.5%) accuracy in distinguishing between *positive-AI* and *negative-AI*, indicating that the manipulation was effective.

To further validate the strength of the manipulation, we conducted a second manipulation check that compared the number of positive and negative strategies present in the generated emails. To closely mirror the study stimuli, we generated 18 emails based on the writing scenarios used in the experiment (see Section 3.2.3). The dataset comprised three scenarios, three conditions (positive politeness, negative politeness, and default GPT-4o), and two emails per scenario–condition combination. Two independent coders applied the codebook to identify all positive and negative strategies in each email. We assessed interrater agreement using the Intraclass Correlation Coefficient (ICC), focusing on ICC(3,k) because the final scores reflect the average of two fixed raters. The resulting ICC(3,k) was 0.88 for negative strategies and 0.77 for positive strategies, both exceeding the threshold for good reliability according to Koo and Li's guidelines [31].

ANOVA results showed a significant effect of AI politeness condition (positive, negative, default) on strategy use for both negative strategies, $F(2, 9) = 163.20$, $p < .001$, and positive strategies, $F(2, 9) = 71.73$, $p < .001$. Post hoc t-tests further clarified these effects. Compared with the default condition, emails generated with negative-strategy prompts contained significantly more negative strategies [$M_{\text{neg}} = 21.75$, $SD = 4.87$ vs. $M_{\text{default}} = 9.50$, $SD = 2.37$; $t = 5.55$, $p < .001$], but showed no significant increase in positive strategies [$M_{\text{pos}} = 4.83$, $SD = 1.51$ vs. $M_{\text{default}} = 4.00$, $SD = 1.61$; $t = 0.93$, $p = .377$]. A parallel pattern was observed for positive-strategy prompts: compared with the default condition, they produced significantly more positive strategies [$M_{\text{pos}} = 11.75$, $SD = 4.19$ vs. $M_{\text{default}} = 4.00$, $SD = 1.61$; $t = 4.23$, $p = .002$], but no significant difference in negative strategies [$M_{\text{neg}} = 10.67$, $SD = 2.21$ vs. $M_{\text{default}} = 9.50$, $SD = 2.37$; $t = 0.88$, $p = .398$]. Together, these findings confirm successful manipulation.

3.2.3 Scenarios. The scenarios used in the study include professional situations that involve making a request. Two sets of scenarios were used in the study — three scenarios were used in writing tasks to guide email writing, and they were slightly modified and mixed with seven additional scenarios to collect participants' social perception before writing tasks. The purpose of using two sets of scenarios was to reduce anchoring effects.

We initially created 18 scenarios adapted from previous work [20], and they broadly fall into three categories: high-risk situations (high D, P, R values), low-risk situations (low D, P, R values), and medium-risk scenarios that are more ambiguous. To validate that the scenarios are perceived as we intended, we performed a pilot study with 18 participants ($N = 18$) who rated all scenarios along the three perception dimensions (D, P, R). The results confirmed that perceptions towards the scenarios generally aligned with our intentions, and we selected three scenarios from the medium-risk category with the most variance across participants to use as writing tasks. We chose these scenarios because they are more open to interpretation, thus there is potentially more room for AI's influence. The three writing scenarios (**WS1–3**) are listed as follows:

WS1: You're enrolled in a course and have interacted with the course TA (teaching assistant) during office hours. You write an email to ask if the TA could clarify a specific concept from a recent class discussion or recommend some additional study materials related to the topic.

WS2: You are working on a project with a colleague, and you're at a stage where you need their part of the work to be completed in order to proceed with your own tasks. You write an email

about their progress and to find out if they can finish their part within the week to keep the project on schedule.

WS3: You're working on a group project for a university course and have had a few discussions with your project advisor. You write an email to ask if the advisor could provide feedback on your current project outline or suggest any resources that might help improve your presentation.

We additionally selected seven scenarios to form the second set of scenarios used to collect pre-writing social perception. Among them, two were selected from the high-risk category, two from the low-risk category, and three from the medium-risk category. The writing scenarios were also slightly modified in the pre-writing survey scenarios. The full list of pre-writing survey scenarios and descriptions is included in Appendix B, and scenarios that correspond to writing scenarios are highlighted.

3.3 Participants

We recruited 66 participants through Prolific. All recruited participants reside in English-speaking countries to ensure that they have adequate language skills to complete the writing tasks in English. We excluded a total of 14 participants who did not have meaningful interaction with the AI writing assistants or make meaningful efforts to complete the tasks, leaving 52 in the final dataset. The filtering criteria were (satisfying any of them): 1) if they did not complete the task; 2) if they did not use any AI generation; 3) if in at least one of the two AI writing tasks they spent less than 2 minutes and wrote less than 10% of words (excluding AI-generated text) in the final email. The final dataset consists of 27 NESs and 25 NNESs, including 28 female participants and 23 male participants, with one participant not disclosing their gender. Participants' ages range from 22 to 66 ($Mean = 39.29, SD = 10.98$).

3.4 Study Procedure

Each participant first completed demographic surveys and then rated their baseline social perceptions of distance (D), power (P), and imposition (R) across 10 randomly ordered scenarios. They then completed the three writing tasks with the three writing scenarios that were selected from the 10 scenarios and slightly modified. The three writing scenarios were presented in random order. Each of the emails was written in one of three conditions: the *baseline* condition, or one of the two AI conditions with AI tools that use positive or negative politeness strategies (*positive-AI* or *negative-AI*).

All participants completed the baseline condition before moving into the AI conditions, and the order of AI conditions was counterbalanced. The baseline condition was completed before the AI-assisted conditions to avoid potential influence after exposure to AI suggestions. After each of the AI-assisted writing tasks, participants were again asked to provide their social perception ratings on distance, power and imposition. These post-writing ratings were compared with pre-writing ratings to capture changes in perception. The study procedure is presented in Figure 2.

Each study session took about 30 minutes in total, and each participant was remunerated £5 for their participation. To encourage high-quality task completion, the top 10% participants with the highest average writing quality were provided a £3 bonus payment. The study protocol was approved by the University of Waterloo Research Ethics Board (REB).

3.5 Measurements

To test the hypotheses, we focused on two main dependent variables — change in the number of politeness strategies (as compared to the baseline condition without AI assistance) and change in social perception. In addition, we included two covariates — reliance on AI and people's baseline

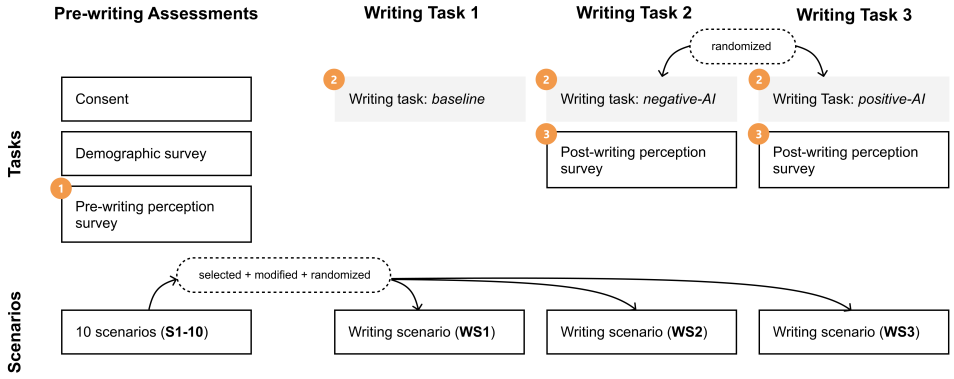


Fig. 2. The study procedure includes three main components: 1) pre-writing survey, 2) writing tasks, and 3) post-writing survey. The pre-writing survey collected perceptions on 10 request scenarios, and each writing task was completed on one of three scenarios selected from the 10. Post-writing survey collected perceptions on the scenario used in the respective writing tasks.

perceptions of social scenarios. We additionally investigated the quality of the emails and how it was affected by AI usage.

3.5.1 Change in Politeness Strategies. To study **H1**, we measured changes in politeness strategies as the dependent variable. For each participant, four different scores were calculated by comparing the number of positive and negative strategies used in the *positive-AI* and *negative-AI* conditions to those in the *baseline* condition. The four scores are defined as follows using mathematical notation:

- Change in positive strategies (*positive-AI* condition):

$$\Delta\text{Positive}_{\text{positive-AI}} = \#\ \text{Positive}_{\text{positive-AI}} - \#\ \text{Positive}_{\text{baseline}}$$

- Change in negative strategies (*positive-AI* condition):

$$\Delta\text{Negative}_{\text{positive-AI}} = \#\ \text{Negative}_{\text{positive-AI}} - \#\ \text{Negative}_{\text{baseline}}$$

- Change in positive strategies (*negative-AI* condition):

$$\Delta\text{Positive}_{\text{negative-AI}} = \#\ \text{Positive}_{\text{negative-AI}} - \#\ \text{Positive}_{\text{baseline}}$$

- Change in negative strategies (*negative-AI* condition):

$$\Delta\text{Negative}_{\text{negative-AI}} = \#\ \text{Negative}_{\text{negative-AI}} - \#\ \text{Negative}_{\text{baseline}}$$

The number of positive and negative strategies was determined through qualitative coding of participants' emails. Each email was manually annotated for instances of positive and negative politeness strategies. The codebook was developed based on Brown & Levinson's politeness theory [11] and was adapted to ensure both comprehensiveness and mutual exclusivity of the codes. The final codebook includes 14 positive politeness strategies and 14 negative politeness strategies.

Using the codebook, two raters, including the first author of the paper, independently coded 20% of emails, which follows conventions in the HCI community [48][26]. To assess interrater reliability, we calculated the Intraclass Correlation Coefficient (ICC), focusing on ICC3 (single fixed raters). For number of negative strategies, the ICC3 was 0.71, indicating moderate reliability. For number of positive strategies, the ICC3 was 0.86, indicating good reliability. The guidelines for interpreting the data were adopted from Koo and Li [31]. These results demonstrate that the interrater reliability

is sufficient to support the use of these ratings in the analysis. The first author of the paper then continued to rate the rest of the email samples, and the analysis was based on data coded by the first author.

3.5.2 Social Perception Change. To study **H2**, the change in social perception was measured by comparing social perception ratings in the post-writing survey and the pre-writing survey (for the scenario that matched the one used in the writing task). The calculation of perception change is represented below:

- Change in Distance perception:

$$\Delta D = D_{\text{post-writing}} - D_{\text{pre-writing}}$$

- Change in Power perception:

$$\Delta P = P_{\text{post-writing}} - P_{\text{pre-writing}}$$

- Change in Imposition perception:

$$\Delta R = R_{\text{post-writing}} - R_{\text{pre-writing}}$$

Social perceptions were measured using a 1–7 self-report Likert scale, with one question assessing each of the dimensions: distance, power, and imposition. Participants were provided with definitions and explanations of these constructs to ensure clarity. The imposition scale was relabeled to use more colloquial terms — "demanding" and "undemanding", which was adapted from [53] to avoid confusion in understanding the questions. On the scales, a score of 1 aligns with low-risk scenarios — it indicates a close relationship, greater power held by the sender, and less demanding requests. A score of 7 aligns with high-risk situations with a more distant relationship, greater power held by the receiver, and more demanding tasks. The full descriptions of the constructs and the scale are included in the Appendix C.

3.5.3 Additional Variables. Our analyses included three additional variables: reliance on AI, individual's baseline social perception, and writing quality of the emails.

Reliance on AI is a mediating variable used to further investigate the effect of AI politeness and language on change in politeness strategies. Reliance on AI is measured by the ratio between the number of accepted AI suggestions and the number of generated AI suggestions. The number of generated AI suggestions includes the number of generation and re-generation requests from the user. While there are other potential ways to reflect reliance (e.g., the proportion of AI-generated text in the final email), this measure was selected because it reflects participants' explicit decisions to accept or reject AI suggestions during the writing process. By contrast, final-text proportion reflects how much AI wording is preserved in the final product, which may not directly indicate reliance, as participants may reject a suggestion but later adopt its idea in their own words, or accept a suggestion but revise it heavily before submission.

Baseline social perception is another covariate used to study how different baseline social perceptions might affect social perception change. Baseline social perception is calculated by taking the average of participants' pre-writing perception ratings across the 10 scenarios on all three dimensions, D, P and R. Variation in this measure reflects individual differences in how people initially interpret and evaluate different social situations.

Email quality is an additional outcome variable included in exploratory analysis. Four raters evaluated the quality of all emails without knowing the conditions the emails were written in. Two of the raters are from the research team, and two raters outside of the research team are the potential receivers of the emails (who have been teaching assistants and instructors for university courses). The raters rated the emails' writing quality on a 1–7 scale based on a rubric that was

adapted from a previous work [28] and [58], which focuses on language accuracy, appropriate email structure including opening and closing, and detailed but succinct delivery of relevant information. The full rubric is included in Appendix D.

We calculated the inter-rater reliability by ICC after rating 10% of emails, and ICC3k (average fixed raters) was (0.80), indicating good reliability between the raters. The raters then independently coded the rest of the emails, and independent reliability was calculated again on the full dataset, with an ICC3k value of (0.78), indicating good reliability. The average rating of the raters was used as the quality rating for analysis.

4 Results

In the analysis, we examined the effects of AI politeness and participant language group on two dependent variables: the number of positive and negative politeness strategies used in the emails (**H1**, **H3**) and participants' social perception ratings (Distance, Power, Imposition) (**H2**, **H3**). First, we summarize the main findings before describing the results in detail.

Strategies used in emails (Section 4.1): Results indicate a significant effect of AI politeness on the use of politeness strategies. There is also a partial effect of language, as *NNESs* adopt more positive strategies than *NESs*, but there is no significant difference in negative strategies. The effect of language is further found to be mediated by participants' reliance on AI tools.

Social perception (Section 4.2): AI politeness has differential effects on the change in distance perception, but changes in power and imposition do not differ across conditions. *negative-AI* changes distance perception more towards the distant end, while *positive-AI* leads to the perception of closer relationships. However, neither *positive-AI* nor *negative-AI* is significantly different from *baseline*, suggesting a potential effect that requires further investigation. *NNESs* and *NESs* are equally impacted.

4.1 Politeness Strategies Used in Emails (H1 Confirmed and H3a Partially Confirmed)

4.1.1 Effects of AI politeness and language on Change in Politeness Strategies. We tested **H1** and **H3a**, which hypothesized that AI politeness (*positive-AI* vs. *negative-AI*) and language (*NNES* vs. *NES*) would both influence changes in politeness strategies people used in the emails. Results fully confirmed **H1** and partially confirmed **H3a**.

The outcome variable, number of politeness strategies included in the emails, is measured by qualitatively coding the emails and counting the number of positive and negative strategies included within them. Here we focused on the change in the number of politeness strategies, which is calculated by subtracting the number of strategies in the *baseline* condition from those in *positive-AI* and *negative-AI* conditions, respectively. The full analysis details were included in Section 3.5.1. We ran a mixed-effects model to analyze the main effects of AI politeness and language, and their interaction on changes in politeness strategies, with individual participants treated as a random effect. In all analyses, the *negative-AI* condition and *NNES* group were used as the reference levels, unless otherwise specified.

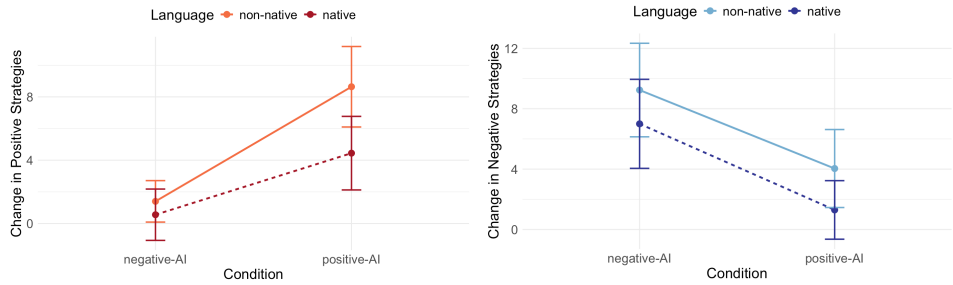
AI politeness has a significant main effect on both the change in positive strategies (ΔPositive) ($\beta = 5.50$, $SE = 0.86$, $t(51) = 6.40$, $p < 0.001$) and the change in negative strategies (ΔNegative) ($\beta = -5.46$, $SE = 1.09$, $t(51) = -5.00$, $p < 0.001$). Figure **??** indicates that the *negative-AI* condition resulted in a greater increase in negative strategies, and the *positive-AI* condition led to a larger increase in positive strategies. This aligns with the expectation and confirms **H1** that **AI politeness changes the corresponding politeness strategies people include in emails.**

There is also a significant main effect of language on ΔPositive ($\beta = -2.52$, $SE = 1.19$, $t(50) = -2.118$, $p = 0.04$) and a close to significant interaction effect between AI politeness and language ($\beta = 3.35$, $SE = 1.67$, $t(50) = 2.004$, $p = 0.051$). In the interaction effect analysis, when using

the *negative-AI* condition as the reference, the main effect of LANGUAGE on ΔPositive was not statistically significant ($\beta = -0.84$, $SE = 1.45$, $t(89.71) = -0.58$, $p = 0.56$). This indicates that *NNESs* and *NESs* do not change their use of positive strategies in different ways influenced by the *negative-AI*. Interestingly, when the reference level switched to the *positive-AI* condition, the effect of LANGUAGE became significant ($\beta = -4.20$, $SE = 1.45$, $t(89.71) = -2.89$, $p < 0.01$). This means that *NNESs* increased their use of positive strategies in the *positive-AI* condition to a higher extent compared to *NESs*, indicating that they are more influenced by AI suggestions.

However, for ΔNegative, no evidence suggests a main effect of LANGUAGE ($\beta = -2.49$, $SE = 1.58$, $t(50) = -1.58$, $p = 0.12$) nor an interaction effect between AI POLITENESS and LANGUAGE ($\beta = -0.50$, $SE = 2.21$, $t(50) = -0.23$, $p = 0.82$). This suggests that *NESs* and *NNESs* are not affected differently in terms of change in negative strategies. The interaction between language and condition for both positive and negative strategies is visualized in Figure 3.

Overall, the results indicate that *NNESs* increase their positive strategies more than *NESs*. Specifically, *NNESs* show a larger increase in positive strategies in the *positive-AI* condition but not in the *negative-AI* condition, as indicated in Figure 3a. This supports that AI's use of positive strategies has a larger impact on *NNESs* than *NESs*. On the other hand, while Figure 3b indicates that *NNESs* exhibit a larger change in negative strategies across conditions, no strong statistical evidence supports this trend. These findings offer partial support for **H3a**, suggesting that **AI politeness strategies only influence *NNESs* more in their use of positive strategies but not their use of negative strategies**.



(a) Positive Strategies (b) Negative Strategies

Fig. 3. The change in use of positive and negative strategies is different across AI POLITENESS and LANGUAGE groups.

4.1.2 Effect on the Human-Written Content in the Emails. To rule out the possibility that the change in email politeness strategies was solely due to the adoption of AI-suggested text that included such strategies, we performed further analysis on the human-written content in the emails. Results indicate that AI POLITENESS also significantly affects the number of positive strategies in the human-written content, further supporting **H1**, while there is only a marginal effect on negative strategies. A marginal interaction effect suggests that AI POLITENESS influences *NNESs* and *NESs* differently on their adoption of positive politeness strategies, providing partial support for **H3a**.

The human-written content was isolated by removing AI-generated content from the final email. AI-generated content was identified by the overlapping words and sentences between AI suggestions and the final email. Using the same qualitative codebook, the total numbers of positive and negative strategies were again quantified from the human-written content. As the human-written content only includes partial emails, calculating changes in politeness strategies by

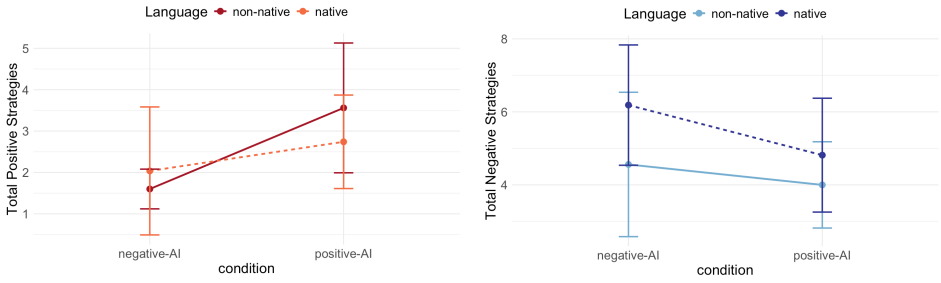
comparing to the baseline condition is no longer methodologically appropriate. Thus, we focus on the total number of positive (#Positive) and negative strategies (#Negative) as outcome variables, instead of changes in number of strategies (ΔPositive and ΔNegative). We then applied a similar mixed-effects model to #Positive and #Negative, treating AI POLITENESS and LANGUAGE as fixed effects, with scenario and individual participants as random effects. The *negative-AI* condition and the *NNES* group were used as the default reference levels, consistent with previous analysis.

The results show a significant main effect of AI POLITENESS on #Positive ($\beta = 1.14$, $SE = 0.43$, $t(49) = 2.69$, $p < 0.01$) and a marginal main effect of AI POLITENESS on #Negative ($\beta = -1.07$, $SE = 0.58$, $t(50) = -1.82$, $p = 0.07$). No significant main effect of LANGUAGE was observed on either #Positive $\beta = 0.00$, $SE = 0.73$, $t(49.33) = 0.01$, $p = 1.00$ or #Negative $\beta = 1.06$, $SE = 0.99$, $t(50.04) = 1.07$, $p = 0.29$.

There is marginal evidence for an interaction effect of AI POLITENESS and LANGUAGE on #Positive ($\beta = 1.58$, $SE = 0.83$, $t(48) = 1.90$, $p = 0.06$). Specifically, when using *NNES* as the reference group, the effect of the AI POLITENESS was significant ($\beta = 1.97$, $SE = 0.60$, $t(48.03) = 3.26$, $p < 0.01$), but when *NES* is treated as the reference group, the effect of AI POLITENESS is not significant ($\beta = 0.39$, $SE = 0.57$, $t(47.65) = 0.69$, $p = .495$) For #Negative, there is no interaction effect between AI POLITENESS and LANGUAGE ($\beta = -0.42$, $SE = 1.19$, $t(48.29) = -0.36$, $p = 0.72$). The data is visualized in Figure 4.

Overall, the results provide additional evidence for **H1** and **H3a**, while also suggesting a need for more nuanced interpretation. They confirm that **AI POLITENESS can shape the language choices in parts written solely by the human**. However, the impact on negative strategies is only marginal, suggesting that the **impact of AI POLITENESS might not be extended to the use of negative strategies in human-written contents.**

The marginal interaction effect between AI POLITENESS and LANGUAGE on the number of positive strategies also aligns with what we found on the full AI-assisted email. Results indicate that **the influence of AI POLITENESS on the use of positive strategies extends to human-written content for *NNES* participants, but not for *NES* participants**. Although the interaction effect is only marginal, this also adds support for **H3a**. On the other hand, aligning with the analysis of the AI-assisted full email, the effect of LANGUAGE on negative strategies is not significant. This means that ***NNES* are not more likely to be influenced by AI POLITENESS on negative strategies**. The evidence offers partial support for **H3a**, highlighting the need for additional research to better understand the effect.



(a) Positive Strategies (b) Negative Strategies

Fig. 4. The use of positive and negative strategies is different across LANGUAGE for human-written content

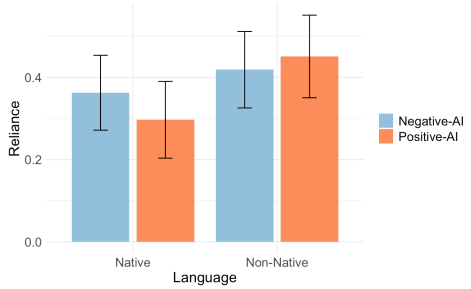


Fig. 5. NNESs are more reliant on AI when it predominantly suggests positive strategies.

4.1.3 Mediating Effect of Reliance on Change in Politeness Strategies. We additionally attempted to uncover the mediating factors that contribute to the influence of AI suggestions on email writing. One factor we hypothesize to play a major role is reliance. We hypothesize that the effect of AI POLITENESS is not mediated by reliance, while the effect of LANGUAGE is mediated by reliance. The data support this hypothesis.

Reliance is measured by calculating the acceptance ratio between the number of accepted AI suggestions and the total number of AI suggestions called. Details of this measurement are included in Section 3.5.3.

We first studied whether reliance is affected by AI POLITENESS and LANGUAGE. We ran a linear mixed-effects analysis on reliance, with AI POLITENESS and LANGUAGE as fixed effects, and participant ID and scenario as random effects. The *negative-AI* condition and the *NNES* group were set as the reference levels. Results indicate no significant main effect of AI POLITENESS ($\beta = -0.02$, $SE = 0.04$, $t(49.61) = -0.42$, $p = 0.68$) but a marginally significant effect of LANGUAGE ($\beta = -0.10$, $SE = 0.06$, $t(49.02) = -1.80$, $p = 0.08$). No significant interaction effect was found between AI POLITENESS and LANGUAGE ($\beta = -0.11$, $SE = 0.08$, $t(48.44) = -1.37$, $p = 0.18$). However, in the *positive-AI* condition, a significant difference in reliance was found between *NNESs* and *NESs*, with *NNESs* being more reliant on AI suggestions ($\beta = -0.15$, $SE = 0.07$, $t(86.10) = -2.26$, $p = 0.03$). There is no significant difference between LANGUAGE groups in the *negative-AI* condition ($\beta = -0.05$, $SE = 0.07$, $t(86.92) = -0.71$, $p = 0.48$). The results are also shown in Figure 5.

Although the interaction between AI POLITENESS and language group was not statistically significant, the effect of LANGUAGE was significant in the *positive-AI* condition but not in the *negative-AI* condition. This suggests that **when AI suggestions predominantly include positive strategies, the *NNES* group is more reliant on AI than the *NES* group**. This is congruent with our other findings, which suggest the effect of LANGUAGE can only be observed in positive strategies but not negative strategies. This provides initial support for our hypothesis that the difference between LANGUAGE groups is mediated by reliance.

To further test our hypothesis, we conducted a Bayesian mixed-effects mediation analysis to examine whether reliance mediated the effect of AI POLITENESS on ΔPositive and ΔNegative. ID and scenario are also treated as random effects. While the primary analyses in this paper used null-hypothesis significance testing, Bayesian mixed-effects approach was used for the mediation analysis due to the limitations of standard mediation packages in handling multilevel models with random effects. We conducted two Bayesian mixed-effects mediation analyses to test whether participants' *reliance* on AI mediated the effect of AI POLITENESS on ΔPositive and ΔNegative, respectively. The results are presented in Table 1.

Table 1. Mediation analysis results by outcome and independent variable

Outcome	Component	Estimate (p-value)
Positive Strategies (AI POLITENESS)	Indirect Effect	-0.12 ($p = .66$)
	Direct Effect	5.23^{***} ($p < .001$)
	Total Effect	5.11^{***} ($p < .001$)
Negative Strategies (AI POLITENESS)	Indirect Effect	-0.15 ($p = .66$)
	Direct Effect	-5.58^{***} ($p < .001$)
	Total Effect	-5.73^{***} ($p < .001$)
Positive Strategies (LANGUAGE)	Indirect Effect	$-0.73^{\cdot}$ ($p = .063$)
	Direct Effect	$-1.64^{\cdot}$ ($p = .092$)
	Total Effect	-2.37^{*} ($p = .022$)
Negative Strategies (LANGUAGE)	Indirect Effect	$-0.89^{\cdot}$ ($p = .063$)
	Direct Effect	-2.24 ($p = .139$)
	Total Effect	-3.13^{*} ($p = .047$)

The results first indicate that **AI POLITENESS has strong direct effects on both positive and negative politeness strategies, with no evidence of mediation through reliance**. This means that simply **being exposed to a certain type of AI suggestion (positive vs. negative) is enough to shape participants' writing**. On the other hand, **the effect of LANGUAGE is mediated by reliance**. Specifically, for positive strategies, LANGUAGE has marginally significant influence on ΔPositive both directly and indirectly, and reliance partially mediates the effect. For ΔNegative, the effect of LANGUAGE is primarily indirect, suggesting that **differential changes in politeness strategies are not fully due to LANGUAGE background but because *NNES* leans on AI more often**.

4.2 Social Perception (H2 partially confirmed, H3b not confirmed)

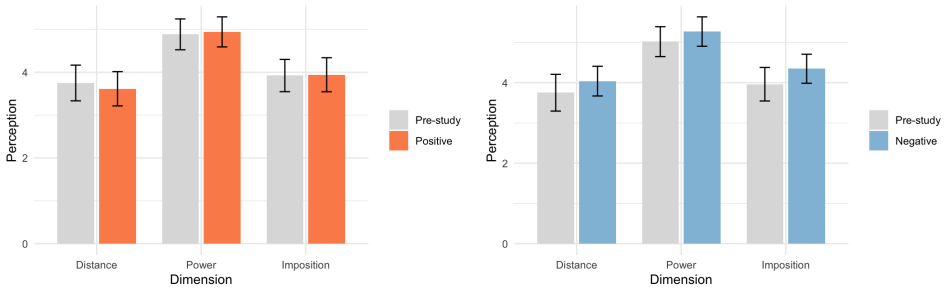
4.2.1 Effect of AI POLITENESS and LANGUAGE on Social Perception. Analysis of the number of politeness strategies included in the emails confirmed that the written content is influenced by AI POLITENESS and the effect is different across LANGUAGE groups. Given the theoretical grounding that politeness strategies used are linked to social perception on three dimensions, distance, power and imposition, we further tested **H2** to see if AI POLITENESS affects social perception. Results partially confirmed this hypothesis, indicating an effect of AI POLITENESS on change in distance perception, but no effect was found on the power and imposition dimension. Additionally, we analyzed if participants in different LANGUAGE groups are impacted differently (**H3b**). This hypothesis is not confirmed, indicating that *NNES* and *NES* are not influenced differently by AI suggestions in terms of social perception.

In the analysis, we also used a mixed-effects model with AI POLITENESS and LANGUAGE as fixed effects, treating scenarios and individual participants as random effects. The outcome variables were changes in perception (ΔD, ΔP, ΔR), calculated as the difference between post-writing perception (after using AI) and pre-writing perception. For each participant, a total of six perception changes were calculated, ΔD, ΔP, ΔR for both *negative-AI* and *positive-AI* conditions. Changes in perception was compared across *negative-AI* and *positive-AI*, with *negative-AI* condition and *NNES* used as reference levels. The results show that the AI POLITENESS had a significant main effect on ΔD ($\beta = -0.61$, $SE = 0.25$, $t(49.67) = -2.42$, $p = 0.02$), with *negative-AI* leading to a larger increase in social distance perception. However, there is no significant effect of AI POLITENESS on ΔP

($\beta = -0.19$, $SE = 0.29$, $t(101) = -0.65$, $p = 0.51$) or ΔR ($\beta = -0.37$, $SE = 0.31$, $t(50.38) = -1.22$, $p = 0.23$). The effect of LANGUAGE is not significant on any of the three dimensions, and there is also no significant interaction effect between AI POLITENESS and LANGUAGE.

This result suggests that the ΔD differs across AI POLITENESS conditions. **In the *negative-AI* condition, participants tend to change their distance perception more towards the distant end, compared to the *positive-AI* condition**. *NNES* and *NES* are not different in the magnitude of perception change. The overall result for all three dimensions is visualized in Figure 6. This aligns with Brown and Levison's theory connecting use of negative strategies with the perception of more distant relationships, however, the connection to the power and imposition dimension is not supported by our results.

While it confirms AI POLITENESS leads to differences in ΔD, we did further analysis to directly compare post-writing perception ($D_{\text{post-writing}}, P_{\text{post-writing}}, R_{\text{post-writing}}$) with pre-writing perception ($D_{\text{pre-writing}}, P_{\text{pre-writing}}, R_{\text{pre-writing}}$) to confirm whether writing with *positive-AI* and *negative-AI* changes perception from pre-writing baseline. The result does not provide strong statistical evidence to suggest a significant change. We did this analysis through a mixed effect model that was run on the perception value, instead of perception change, with a new variable TIME to capture the perception ratings collected pre-writing task vs. post-writing task. On the distance dimension ($D_{\text{pre-writing}}$ vs. $D_{\text{post-writing}}$), we did not observe an interaction effect between TIME and AI POLITENESS ($\beta = 0.42$, $SE = 0.31$, $t(151) = 1.37$, $p = 0.17$), and when post-writing is set as the reference, there is no main effect of AI POLITENESS ($\beta = -0.30$, $SE = 0.22$, $t(152) = -1.37$, $p = 0.174$). This shows that distance perception did not change in different directions after writing with AI tools. This contradicts findings on the effect of AI POLITENESS on ΔD, which calls for more careful interpretation of the results. It can suggest that **the effect of the AI POLITENESS on perception is very subtle**, and is hard to detect through a one-off experiment like the ones we conducted. Future analysis with a larger sample size and long-term exposure to AI manipulation is needed to find more robust results.



(a) Positive-AI (b) Negative-AI

Fig. 6. Social perception comparison before and after writing tasks.

4.2.2 Mediating Effect of Change in Politeness Strategies on Change in Distance Perception. As the previous section suggests an effect of AI POLITENESS on change in politeness strategies and change in perception, we hypothesize that the effect on perception could either come from a direct effect of exposure to AI suggestions, or it could be mediated by the change in politeness strategies in writing (i.e. people's perception changes as a result of the changes in their writing). Through our analysis, we were able to confirm that the change in perception can be attributed to the direct effect of the AI POLITENESS, not mediated by the change in writing.

We conducted a mediation analysis using a mixed-effects model that included random intercepts for scenario and participant ID. AI POLITENESS served as the predictor, and the change in politeness strategies was specified as the mediator. As changes in politeness strategies include both ΔPositive and ΔNegative, we employed changes in relevant strategies as our measurement. When the condition is *negative-AI*, AI suggestions predominantly include negative strategies, and the ΔNegative is the relevant metric that matters most. Similarly, when the condition is *positive-AI*, the ΔNegative is used to capture the mediating factor.

Results reveal a significant total effect of AI POLITENESS ($\beta = -0.61$, $p = 0.02$), which is aligned with previous findings. There is a significant direct effect ($\beta = -0.64$, $p = 0.02$) and non-significant in-direct effect ($\beta = 0.03$, $p = 0.47$). This indicates that the effect of AI POLITENESS on ΔD is direct, suggesting that **participants' perceptions shifted simply through exposure to AI suggestions, rather than as a result of changes in their use of politeness strategies during writing**.

4.3 Additional Analysis

For more in-depth analysis, we ran two additional exploratory analyses. In the first analysis, we study the effect on email quality, evaluated by external raters. In the second analysis, we study whether people with different baseline social perceptions would be affected differently in terms of their social perception change. The analysis process and results are presented in the following.

4.3.1 Effect of AI POLITENESS and LANGUAGE on email quality evaluations. In addition to the effect of AI politeness strategies on participants' writing strategies and social perception, we tested whether using AI that adopts these strategies might affect how the emails are perceived, especially the writing quality. The results suggest that using AI tools leads to the emails being perceived more negatively, but the effect does not differ across AI POLITENESS and LANGUAGE.

We performed a mixed-effects analysis to examine the main effects of writing conditions (baseline vs. *positive-AI* vs. *negative-AI*) and LANGUAGE (*NNES* vs. *NES*) on email quality, as well as their interaction, with individual participants and writing scenarios as the random effect. **H1** and **H2** ask about the differential impact of *positive-AI* versus *negative-AI*, so we focus on measurements of change (ΔPositive, ΔNegative, ΔD, ΔP, ΔR). In the analysis of email quality, we focus on the decrease in quality in both conditions, so we study the quality rating itself. Perceived writing quality was measured by asking 4 raters who were blind to the conditions to rate the emails, and the final quality rating is the average across the raters. Details of the measurement is included in Section 3.5.3.

The results revealed a main effect of writing condition on email quality, with significant difference between *negative-AI* and *baseline* condition ($\beta = -0.39$, $SE = 0.13$, $t(100.39) = -3.05$, $p = 0.003$), and weak statistical evidence suggesting difference between *positive-AI* and *baseline* condition ($\beta = -0.22$, $SE = 0.13$, $t(100.27) = -1.75$, $p = 0.08$), which indicates a decrease in evaluated email quality after writing with AI. However, the difference between the two AI conditions (*negative-AI* vs. *positive-AI*) is not significant ($\beta = -0.17$, $SE = 0.13$, $t(100.50) = -1.30$, $p = 0.20$). We also did not find a main effect of LANGUAGE or an interaction effect between writing condition and LANGUAGE, suggesting that *NNESs* and *NESs* do not differ in their email quality and NNESs' email quality is not more impacted by AI suggestions.

4.3.2 The Effect of Baseline Social Perception on the Change in Perception. WWhile previous analyses show that LANGUAGE group does not predict the changes in social perception, we ran an additional analysis to test whether another individual factor — baseline social perception — affects social perception change. Baseline social perception was measured by averaging each participant's pre-writing ratings across 10 scenarios and was used as an indicator of their trait-level

tendency to perceive social situations. We are interested in this factor as we hypothesize that people whose baseline perception is closer to the population average may be more susceptible to the influence of AI POLITENESS, as they have less strongly defined views about social situations and thus are more likely to be impacted by AI. However, results suggest a trend opposite to the initial hypothesis, showing a convergence effect in perception. Participants with more extreme baseline social perceptions were more influenced by AI suggestions in terms of perception change, and this effect applies across three dimensions.

The effect of baseline social perception is also tested with a mixed-effects model, with ΔD, ΔP, ΔR as outcome variables, baseline social perception as fixed effect and ID and scenario as random effects. As indicated in Table 2, change in perception is negatively correlated with baseline social perception in all three dimensions. **Participants with initial low perception value tend to increase their perception, and those with initial high value tend to decrease their perception.** This can lead to their final perception being more homogenized after writing with AI.

As an additional exploratory analysis, the convergence effect we found might suggest that AI reduced individual differences among participants in terms of social perception, leading to more similar views of the social situation. However, the result should also be interpreted with caution, as we are not able to compare changes in perception in the *baseline* condition where users did not write with AI tools. Thus, we can not rule out the possibility that the convergence is the result of writing. Our results can only suggest preliminary directions that require further investigation.

Table 2. Change in perception is negatively correlated with baseline social perceptions

Predictor	Estimate	Std. Error	T-Value	P-Value
Baseline Distance Perception	-0.60	0.22	-2.68	<0.01**
Baseline Power Perception	-1.49	0.38	-3.90	<0.01**
Baseline Imposition Perception	-0.87	0.25	-3.47	<0.01**

5 Discussion

5.1 Summary of Results

The results confirmed our hypothesis that the politeness strategies included in AI suggestions influence people's use of politeness strategies in email writing and shape their social distance perception. The effect on the usage of politeness strategies is prominent and can be extended to the human-written components in the emails. The effect on social perception, however, is more subtle and may require nuanced interpretation and additional studies. We also found that changes in politeness strategies and social perception were correlated with language background. While the NNES group was more influenced by AI suggestions in terms of their use of politeness strategies, their social perception was not impacted differently. Mediation analysis suggests that AI's influence on their writing is partly mediated by reliance, suggesting that NNESs are more influenced by AI because they rely on the tools more than NESs. Additionally, we found that using AI, regardless of politeness suggestion type, decreased the quality of the emails. Also, we found a convergence effect, where people with initial extreme social perceptions tend to converge their perceptions towards the middle after writing tasks using AI assistants.

In the following, we discuss potential reasons behind AI's influence on writing and social perception, and why NNESs are more influenced by AI, connecting with results in our mediation analysis and drawing on relevant literature. Also, we will provide some design implications as

insights to guide the building of AI systems that are more aware of social and cultural context, and are able to facilitate social interactions instead of hindering them.

5.2 Exploring Mechanisms behind AI's Influence on Writing and Social Perception, and the Added Impact on NNESs

5.2.1 Effect of AI Politeness on Politeness Strategies in Emails and Social Perception. Results show a significant effect of AI Politeness on politeness strategies in emails, and this also extends to the human-written components. This could be an influence directly applied at the linguistic level, or potentially come from the shaping of social perceptions and then indirectly change writing strategies. While both theories are supported by existing literature and align with aspects of our results, neither fully accounts for the underlying process.

Previous theories and studies have explored the effect of linguistic alignment. Based on the *interactive alignment theory* [45], during conversations, people would automatically align mental models, which leads to synchronization in linguistic representations. This is supported by many other works that suggest people's linguistic style converges during the establishment of friendships [32] and in online communities like Reddit [3]. In addition to linguistic alignment in human-human interaction, this alignment effect has also been found in human-AI interactions, especially in interaction with conversational agents [55]. We believe that the automatic process of aligning mental models and converging linguistic style might be activated when human writers are exposed to AI suggestions in our study. Our results provide additional evidence to suggest that the linguistic alignment effect also applies to human-AI interactions.

On the other hand, our results also indicate that changes in social perception happen concurrently with changes in writing strategies. This suggests that another potential mechanism behind the change in politeness strategies in writing might be through the change in social perception. However, the mixed findings in our study suggest that while AI politeness might have an impact on social perception, the effect is relatively subtle, and is not the main reason that contributes to changes in politeness strategies used in the emails.

We additionally hypothesize that there might be a bottom-up effect where people's writing changes as a result of using AI assistants, and thus changes their social perception. However, mediation analysis does not support this relationship. Thus, there are reasons to believe that the link between change in writing strategies and change in perception is relatively weak, and AI suggestions impacted change in these two variables through different mechanisms. There is room to explore further the specific reasons behind the change, to help provide more implications for future design and development of AI tools.

5.2.2 Effect of language on Politeness Strategies in Emails. Results found NNESs are more affected by AI politeness in terms of their strategy change in emails. This effect is partly mediated by reliance on AI tools.

This aligns with previous work on the usage of AI writing assistants among NNESs, suggesting that NNESs are generally more reliant on AI tools and believe that these tools are better than their judgments. This reliance is also found to be more prevalent among people with lower proficiency levels than those with higher levels [30].

Additionally, we found the differential effect of language is mostly significant with positive strategies but not negative strategies. We propose two possible explanations for this. Firstly, by definition in the Brown & Levinson theory, negative strategies are considered to be closer to what we consider "conventionally polite". This means that this could be a relatively more familiar type of politeness expression to many people. Given this assumption, both NESs and NNESs included in the study might have built enough familiarity towards negative politeness strategies, but positive

politeness strategies can be more foreign to NNESs because they are less frequently used in general. When presented with such AI suggestions, this lack of familiarity might lead to more re-strategizing in email writing for NNESs. Secondly, the AI suggestions, although carefully fine-tuned, might still include negative strategies in the *positive-AI* condition. Our manipulation check (Section 3.2.2) shows that people are more likely to mistake *negative-AI* for default GPT suggestions, while *positive-AI* is more distinguishably different from default GPT suggestions. With negative strategies present in both *positive-AI* and *negative-AI* conditions, the effect of AI politeness might not be as salient, thus not reflected in our statistical test results.

5.3 Design Implications

These negative consequences may arise from AI writing assistants being insensitive to social context. For example, in our study, the AI assistants were calibrated to only employ positive or negative strategies regardless of the social scenario. This insensitivity to social context is also prevalent in popular AI writing assistant tools including ChatGPT [42], Grammarly [23], and Quillbot [47]. Thus, we propose the following design implications to support context awareness in AI-mediated communication while preserving the writer's unique personal characteristics in the process. This will also benefit NNESs in better evaluating the accuracy and appropriateness of AI suggestions.

5.3.1 Design Implication 1: Incorporating Social-Context Awareness in AI-Mediated Communication. In our study, the AI writing assistants were calibrated to predominantly incorporate positive or negative strategies. When AI suggestions are misaligned with writers' initial perceptions of the social situation, it may create distorted perceptions with inappropriate writing strategies. This calls for the design of AI writing assistants that are more aware of the social context, and suggest content suitable for the social scenario.

The relevant social context is both subjective and objective. AI writing assistants can be designed to capture the objective social relationship and provide suggestions that fit the general context of a specific type of writing. For example, IntroAssist [22] is designed specifically to help write introductory requests, where the social relationship is relatively consistent. However, social perception can also be subjective and dynamic, as the relationship is negotiated between members in conversation [12]. Therefore, AI writing assistants should offer suggestions that align with the writer's interpretation of the scenario while ensuring the message is accurately conveyed to the receiver. Previous studies have explored the development of receiver personas and providing tailored feedback based on the intended persona [7]. Additionally, AI assistants can also actively collect the writer's perception in the social scenario, to adapt writing strategies accordingly. For example, asking writers to input their perception of distance, power and imposition can help AI suggest politeness strategies that are better aligned with the writer's evaluation. Politeness is one of the important aspects of AI-mediated communication, while other underexplored aspects might also be relevant. Further studies are needed to explore the link between social context and language use, to create AI suggestions that better align with the writer's intentions.

5.3.2 Design Implication 2: Preserving Writer's Individualized Social Perception and Writing Style. Additional analysis in the study suggests a convergence effect in perception, indicating that AI suggestions lead to more homogenized social perceptions. This homogenizing effect aligns with previous work that suggests AI might homogenize cultural differences [1] or lead to less variation in creative ideations [4]. This finding can have significant implications. While having less diverse social perception may facilitate communication, as the reduction in perception differences might bridge misunderstandings, it could also have negative implications by diminishing individual characteristics.

This suggests design implications for AI writing assistants to help writers preserve their uniqueness in the communication process. This can be addressed by developing AI assistants that mimic the writing styles of writers to minimize discrepancies between AI writing style and human writing style. Also, as the results suggest that AI assistants' influence on people's perception and writing strategies is largely unconscious, it is important to develop mechanisms to promote awareness among writers and reflect on how they have been influenced by AI. There is potential to leverage visualizations to help writers identify the discrepancies between AI suggestion style and writers' personal style.

5.3.3 Support for non-native English speakers. As indicated in Section 4.1.3, NNESs' writing strategies are more influenced by AI writing assistants, and this is mediated by their higher reliance on AI tools. This might be due to the lack of confidence in their ability to write socially appropriate messages and their tendency to treat AI as a tool with authority. Previous studies have suggested that the main challenge NNESs face when interacting with AI assistants is the lack of ability to assess the appropriateness of AI suggestions [30]. There are multiple layers of challenges, including ensuring language accuracy, delivering messages that reflect the writer's intentions, and adopting communication strategies that align with social norms [13]. Thus, to help NNESs better evaluate AI suggestions, AI assistants should focus on ensuring that the suggestions align with the users' social intentions and are interpreted by the receivers as intended. Previous studies have explored ways to help NNESs navigate the challenges, for example, providing examples to help them evaluate the suggestions in context [30]. We suggest that the previous two design implications – the incorporation of contextual awareness and mechanisms to preserve the writer's personal characteristics – would greatly benefit non-native speakers. By emphasizing the social context, NNESs can gain confidence that AI suggestions are tailored to their specific communication scenarios, which allows them to write messages that accurately convey their intentions. Additionally, mechanisms that preserve the writer's unique characteristics will ensure that NNESs are not overly influenced by AI tools predominantly trained on Western-centric data.

In addition, it is important for AI assistants to empower NNESs and not exacerbate their overreliance on AI tools. While tools that are perfectly tailored to the social context and NNESs' personal characteristics can be extremely supportive, they may not directly address their feelings of being foreign or inadequate. On top of providing writers with the most appropriate response based on AI's assessment of the situation, AI assistants for non-native speakers need to focus on scaffolding the process, helping them translate the social context to a more familiar context they could understand, and helping them evaluate the accuracy of their responses using reference points that they can understand from their own culture.

5.4 Limitations and Future Work

We acknowledge that our study design has some limitations. First, we did not directly compare treatment conditions (*positive-AI* and *negative-AI*) with the *baseline* condition where participants wrote without AI. Instead, we examined changes in politeness strategies and social perception across *positive-AI* and *negative-AI*, with the changes calculated relative to the *baseline* condition. Thus, our findings should be interpreted as evidence of the relative influence of different AI suggestion styles, rather than evidence of the overall effect of AI assistance itself. In other words, we do not draw conclusions about whether using AI, compared with not using AI, changes people's writing style or social perceptions.

There is also the limitation of including negative strategies in the *positive-AI* condition. This is due to the limitation of the AI models used in the study, as the default language style of the GPT-4o model we used leans towards the inclusion of many negative strategies, and they are also prevalent

even when the model is instructed to predominantly include positive strategies. The inclusion of negative strategies in the *positive-AI* condition impacted participants' language usage, as their use of negative strategies also increased in the *positive-AI* condition compared to *baseline*. Although this does not impact the validity of the differential effect we found on change in politeness strategies, it might be one of the contributing factors to not finding strong enough evidence in the change in social perception, as the language cues exposed to participants were not different enough across conditions.

Moreover, the lack of significant change in perception might also be due to the duration of the study, which was designed for short-term interactions rather than focusing on the long-term effects. Similar to what we found on the change in social perceptions, previous studies have also suggested that AI suggestions would have a distortion effect on social perceptions [15], but also suggest that this might happen over long-term usage of AI tools. We assume that the effect on perception might be more long-term than short-term, which can be further studied in future research. It would also be valuable for future longitudinal studies to examine whether changes in writing behaviours and perceptions persist even in the absence of AI assistance.

We also acknowledge a limitation in our sampling approach. In our study, we only recruited participants who resided in English-speaking countries to ensure that all participants had adequate skills to complete the writing tasks in English. This means that all non-native English speakers included in our study were relatively more proficient in English, as they needed to use this language frequently in daily interactions. This may be another limitation of the study, as we could not include participants who may struggle to understand the context of communication and express themselves. In our study, we were able to identify a difference in the use of politeness strategies between language groups, but were not able to establish a difference in perception. Further studies that include participants with diverse language levels can potentially reveal more insights.

Additionally, in our study, we were limited to using simulated scenarios instead of real-world scenarios, because the scenarios needed to be comparable across participants to achieve internal validity. However, with this simulated design, participants might only have a relatively ambiguous image in their minds in terms of the social situation they are reacting to and the person they are writing to. This could potentially influence their writing strategies, as there are no real social consequences of their email writing. We tried to mitigate this effect by giving additional incentives to people with higher email quality, but this may not fully simulate real-life situations. Also, there could be an impact on social perception, as in more ambiguous scenarios, participants' perception might not be accurate, thus the changes in perception may also not reflect reactions in real-life scenarios. The inclusion of scenarios in the AI prompt might also limit the real-world applications of the study. While we included scenario prompts in the study to control for individual variances in prompting skills, real-world writing assistants might not perform as well as the system used in the study. Future work with long-term observations of people in their real-life work settings would provide stronger evidence for the phenomena studied in this paper.

Finally, we acknowledge several limitations in how our findings should be interpreted. Our study focuses on email continuation tasks with scenarios that are neither extremely high-stakes nor low-stakes, which limits generalizability to other writing contexts. Different interface affordances, such as short auto-replies or full-paragraph rewriting, may lead to different patterns of AI use and influence than those observed here. Effects may also differ across interaction contexts, such as informal communication between friends or high-stakes professional situations that involve multiple revision cycles and organizational conventions. As stakes vary, writers may over-rely on AI when they trust its judgment, or under-rely when they feel greater personal responsibility for the message. In addition, our study places primary emphasis on writers' experiences. Although we examined recipients' evaluations of writing quality, other important aspects of perception,

including appropriateness and conveyed social distance, power, and imposition, remain unexplored. These limitations point to the need for future work that examines a broader range of writing modalities, scenarios, and recipient-centred outcomes.

6 Conclusion

In this paper, we studied how AI's incorporation of two types of politeness strategies—negative and positive strategies—impacts people's writing strategies and social perception. The results suggest that writers adopt the politeness strategies used by AI, and their perception of social distance is shifted in the process. Non-native English speakers are sometimes affected to a greater extent than native English speakers. The results suggest design implications that emphasize social context awareness in AI writing assistants, along with mechanisms that preserve the personal characteristics of writers. Furthermore, tailored and more empowering designs should be developed to better support the non-native speaker demographic.

References

[1] Dhruv Agarwal, Mor Naaman, and Aditya Vashistha. 2025. AI Suggestions Homogenize Writing Toward Western Styles and Diminish Cultural Nuances. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 1117, 21 pages. doi:10.1145/3706598.3713564

[2] Mike Allen. 2017. *The SAGE Encyclopedia of Communication Research Methods*. SAGE Publications, Inc., Thousand Oaks, CA, USA. doi:10.4135/9781483381411

[3] Aparna Ananthasubramaniam, Hong Chen, Jason Yan, Kenan Alkiek, Jiaxin Pei, Agrima Seth, Lavinia Dunagan, Minje Choi, Benjamin Litterer, and David Jurgens. 2023. Exploring Linguistic Style Matching in Online Communities: The Role of Social Context and Conversation Dynamics. In *Proceedings of the First Workshop on Social Influence in Conversations (SICon 2023)*, Kushal Chawla and Weiyan Shi (Eds.). Association for Computational Linguistics, Toronto, Canada, 64–74. doi:10.18653/v1/2023.sicon-1.7

[4] Barrett R Anderson, Jash Hemant Shah, and Max Kreminski. 2024. Homogenization Effects of Large Language Models on Human Creative Ideation. In *Proceedings of the 16th Conference on Creativity & Cognition* (Chicago, IL, USA) *(C&C '24)*. Association for Computing Machinery, New York, NY, USA, 413–425. doi:10.1145/3635636.3656204

[5] Kenneth C. Arnold, Krysta Chauncey, and Krzysztof Z. Gajos. 2018. Sentiment Bias in Predictive Text Recommendations Results in Biased Writing. In *Proceedings of the 44th Graphics Interface Conference* (Toronto, Canada) *(GI '18)*. Canadian Human-Computer Communications Society, Waterloo, CAN, 42–49. doi:10.20380/GI2018.07

[6] Kathleen Bardovi-Harlig and Beverly S Hartford. 1990. Congruence in native and nonnative conversations: Status balance in the academic advising session. *Language learning* 40, 4 (1990), 467–501.

[7] Karim Benharrak, Tim Zindulka, Florian Lehmann, Hendrik Heuer, and Daniel Buschek. 2024. Writer-Defined AI Personas for On-Demand Feedback Generation. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 1049, 18 pages. doi:10.1145/3613904.3642406

[8] Diane S. Berry, James W. Pennebaker, Jennifer S. Mueller, and Wendy S. Hiller. 1997. Linguistic Bases of Social Perception. *Personality and Social Psychology Bulletin* 23, 5 (1997), 526–537. arXiv:https://doi.org/10.1177/0146167297235008 doi:10.1177/0146167297235008

[9] Sigrun Biesenbach-Lucas. 2007. Students Writing E-Mails to Faculty: An Examination of E-Politeness among Native and Non-Native Speakers of English. *Language Learning & Technology* 11 (2007), 59–81.

[10] Jessica Y Bo, Sophia Wan, and Ashton Anderson. 2025. To Rely or Not to Rely? Evaluating Interventions for Appropriate Reliance on Large Language Models. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 905, 23 pages. doi:10.1145/3706598.3714097

[11] Penelope Brown and Stephen C. Levinson. 1987. *Politeness: Some Universals in Language Usage*. Vol. 4. Cambridge University Press, Cambridge, UK.

[12] Herbert H. Clark and Susan E. Brennan. 1991. Grounding in Communication. In *Perspectives on Socially Shared Cognition*, Lauren B. Resnick, John M. Levine, and Stephanie D. Teasley (Eds.). American Psychological Association, 127–149. doi:10.1037/10096-006

[13] Maria Economidou-Kogetsidis. 2011. "Please answer me as soon as possible": Pragmatic failure in non-native speakers' e-mail requests to faculty. *Journal of Pragmatics* 43, 13 (2011), 3193–3215.

[14] Liye Fu, Susan Fussell, and Cristian Danescu-Niculescu-Mizil. 2020. Facilitating the Communication of Politeness through Fine-Grained Paraphrasing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language*

Processing (EMNLP), Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu (Eds.). Association for Computational Linguistics, Online, 5127–5140. doi:10.18653/v1/2020.emnlp-main.416

[15] Yue Fu, Sami Foell, Xuhai Xu, and Alexis Hiniker. 2024. From Text to Self: Users' Perception of AIMC Tools on Interpersonal Communication and Self. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 977, 17 pages. doi:10.1145/3613904.3641955

[16] Erving Goffman. 1967. Interaction ritual. Essays on face-to-face interaction.

[17] Sage L. Graham and Claire Hardaker. 2017. *(Im)politeness in Digital Communication.* Palgrave Macmillan UK, London, 785–814. doi:10.1057/978-1-137-37508-7_30

[18] 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.

[19] Chien-Ju Ho, Aleksandrs Slivkins, Siddharth Suri, and Jennifer Wortman Vaughan. 2015. Incentivizing High Quality Crowdwork. In *Proceedings of the 24th International Conference on World Wide Web* (Florence, Italy) *(WWW '15)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 419–429. doi:10.1145/2736277.2741102

[20] Peter J. Hofmann. 2003. *Language Politeness: Directive Speech Acts in Brazilian Portuguese, Costa Rican Spanish, and Canadian English.* D.A. thesis. State University of New York at Stony Brook, Stony Brook, NY, USA.

[21] Jess Hohenstein, Rene F Kizilcec, Dominic DiFranzo, Zhila Aghajari, Hannah Mieczkowski, Karen Levy, Mor Naaman, Jeffrey Hancock, and Malte F Jung. 2023. Artificial intelligence in communication impacts language and social relationships. *Scientific Reports* 13, 1 (2023), 5487.

[22] Julie S. Hui, Darren Gergle, and Elizabeth M. Gerber. 2018. IntroAssist: A Tool to Support Writing Introductory Help Requests. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3173574.3173596

[23] Grammarly Inc. 2025. *Grammarly.* https://www.grammarly.com

[24] Maurice Jakesch, Advait Bhat, Daniel Buschek, Lior Zalmanson, and Mor Naaman. 2023. Co-Writing with Opinionated Language Models Affects Users' Views. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 111, 15 pages. doi:10.1145/3544548.3581196

[25] 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* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3290605.3300469

[26] Janet G Johnson, Danilo Gasques, Tommy Sharkey, Evan Schmitz, and Nadir Weibel. 2021. Do You Really Need to Know Where "That" Is? Enhancing Support for Referencing in Collaborative Mixed Reality Environments. 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 514, 14 pages. doi:10.1145/3411764.3445246

[27] Kowe Kadoma, Marianne Aubin Le Quere, Xiyu Jenny Fu, Christin Munsch, Danaë Metaxa, and Mor Naaman. 2024. The Role of Inclusion, Control, and Ownership in Workplace AI-Mediated Communication. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 1016, 10 pages. doi:10.1145/3613904.3642650

[28] Stefan D Keller, Ruth Trüb, Emily Raubach, Jennifer Meyer, Thorben Jansen, and Johanna Fleckenstein. 2023. Designing and validating an assessment rubric for writing emails in English as a foreign language. *Research in Subject-matter Teaching and Learning (RISTAL)* 6, 1 (2023), 16–48.

[29] Margaret L Kern, Johannes C Eichstaedt, H Andrew Schwartz, Lukasz Dziurzynski, Lyle H Ungar, David J Stillwell, Michal Kosinski, Stephanie M Ramones, and Martin EP Seligman. 2014. The online social self: An open vocabulary approach to personality. *Assessment* 21, 2 (2014), 158–169.

[30] Yewon Kim, Thanh-Long V. Le, Donghwi Kim, Mina Lee, and Sung-Ju Lee. 2025. Design Opportunities for Explainable AI Paraphrasing Tools: A User Study with Non-native English Speakers. In *Proceedings of the 2025 ACM Designing Interactive Systems Conference (DIS '25)*. Association for Computing Machinery, New York, NY, USA, 1061–1083. doi:10.1145/3715336.3735740

[31] Terry K Koo and Mae Y Li. 2016. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine* 15, 2 (2016), 155–163.

[32] Balazs Kovacs and Adam M Kleinbaum. 2020. Language-style similarity and social networks. *Psychological science* 31, 2 (2020), 202–213.

[33] Aikaterini Leontaridou. 2015. *Power and politeness in email communication in the workplace.* Master's thesis. Leiden University.

[34] Stephen C Levinson. 1983. *Pragmatics.* Cambridge University Press, Cambridge, UK.

[35] Hajin Lim, Dan Cosley, and Susan R. Fussell. 2022. Understanding Cross-lingual Pragmatic Misunderstandings in Email Communication. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW1, Article 129 (apr 2022), 32 pages. doi:10.1145/3512976

[36] I. Scott MacKenzie. 2024. *Human-computer interaction: An empirical research perspective* (2nd ed.). Morgan Kaufmann (an imprint of Elsevier), Cambridge, MA. doi:10.1016/C2022-0-02755-0

[37] 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.

[38] Hannah Mieczkowski, Jeffrey T. Hancock, Mor Naaman, Malte Jung, and Jess Hohenstein. 2021. AI-Mediated Communication: Language Use and Interpersonal Effects in a Referential Communication Task. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 17 (apr 2021), 14 pages. doi:10.1145/3449091

[39] Hannah Nicole Mieczkowski and Stacey F., Bent. 2022. *AI-Mediated Communication: Examining Agency, Ownership, Expertise, and Roles of AI Systems.* Ph. D. Dissertation. Stanford University, Stanford, CA, USA. Advisor(s) Jeremy, Bailenson, and Mor, Naaman, and Byron, Reeves,. AAI29756350.

[40] Stephanie Milani, Arthur Juliani, Ida Momennejad, Raluca Georgescu, Jaroslaw Rzepecki, Alison Shaw, Gavin Costello, Fei Fang, Sam Devlin, and Katja Hofmann. 2023. Navigates Like Me: Understanding How People Evaluate Human-Like AI in Video Games. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 572, 18 pages. doi:10.1145/3544548.3581348

[41] David A. Morand and Rosalie J. Ocker. 2003. Politeness Theory and Computer-Mediated Communication: A Sociolinguistic Approach to Analyzing Relational Messages. In *Proceedings of the 36th Annual Hawaii International Conference on System Sciences*. IEEE Computer Society, Big Island, HI, USA, 10 pages. doi:10.1109/HICSS.2003.1173660

[42] OpenAI. 2025. *ChatGPT*. https://openai.com

[43] James W Pennebaker and Laura A King. 1999. Linguistic Styles: Language Use as an Individual Difference. *Journal of Personality and Social Psychology* 77, 6 (1999), 1296.

[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] Martin J Pickering and Simon Garrod. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences* 27, 2 (2004), 169–190.

[46] Ritika Poddar, Rashmi Sinha, Mor Naaman, and Maurice Jakesch. 2023. AI Writing Assistants Influence Topic Choice in Self-Presentation. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 29, 6 pages. doi:10.1145/3544549.3585893

[47] Quillbot. 2025. *Quillbot*. https://quillbot.com

[48] Iulian Radu and Bertrand Schneider. 2019. What Can We Learn from Augmented Reality (AR)? Benefits and Drawbacks of AR for Inquiry-based Learning of Physics. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3290605.3300774

[49] 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. doi:10.1145/3411764.3445557

[50] William R. Shadish, Thomas D. Cook, and Donald Thomas Campbell. 2002. *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin, Boston.

[51] John J Shaughnessy, Eugene B Zechmeister, and Jeanne S Zechmeister. 2000. *Research methods in psychology*. McGraw-Hill, Boston, MA, USA.

[52] Julia Simkus. 2023. *Within-Subjects Design: Examples, Pros & Cons*. Simply Psychology. Accessed: 2026-01-13. https://www.simplypsychology.org/within-subjects-design.html

[53] Bruno Staszkiewicz Garcia. 2018. *The importance of Power, Distance, and Imposition on Spanish verb forms in requests*. Master's thesis. Purdue University, West Lafayette, IN, USA. Open Access Theses, 1457. https://docs.lib.purdue.edu/open_access_theses/1457

[54] Jenny Thomas. 1983. Cross-Cultural Pragmatic Failure. *Applied Linguistics* 4, 2 (07 1983), 91–112. arXiv:https://academic.oup.com/applij/article-pdf/4/2/91/9741677/91.pdf doi:10.1093/applin/4.2.91

[55] Antonia Tolzin and Andreas Janson. 2023. Mechanisms of Common Ground in Human-Agent Interaction: A Systematic Review of Conversational Agent Research. In *Proceedings of the 56th Hawaii International Conference on System Sciences*. Hawaii International Conference on System Sciences, Maui, Hawaii, USA, 342–351. doi:10.24251/HICSS.2023.042

[56] Xiaoxu Yang, Jue Hou, and Zachary W Arth. 2021. Communicating in a proper way: How people from high-/low-context culture choose their media for communication. *International Communication Gazette* 83, 3 (2021), 238–259.

[57] Michael Yeomans, Alejandro Kantor, and Dustin Tingley. 2018. The politeness Package: Detecting Politeness in Natural Language. *The R Journal* 10, 2 (2018), 489–502. doi:10.32614/RJ-2018-079

[58] Soobin Yim, Dakuo Wang, Judith Olson, Viet Vu, and Mark Warschauer. 2017. Synchronous Collaborative Writing in the Classroom: Undergraduates' Collaboration Practices and their Impact on Writing Style, Quality, and Quantity. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, Oregon, USA) *(CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 468–479. doi:10.1145/2998181.2998356

A Positive and negative politeness strategies

Positive Politeness Strategies	
Strategy and Description	**Examples**
Strategy 1: Notice hearer's admirable qualities or possessions, show interest, exaggerate	*'marvelous, fantastic, extraordinary, wonderful, delightful, ravishing, divine, incredible, appalling, ghastly, devastating, outrageous, despicable, revolting, ridiculous, incredible, absolutely, completely'* *"Hey love your new Palm-pilot, can I borrow it sometime?"* *"How absolutely marvelous/extraordinary/devastating/incredible!"*
Strategy 2: Employ phonological slurring to convey in-group membership	*"Heya, gimme a hand with this willya?"*
Strategy 3: Use colloquialisms or slang to convey in-group membership	*"Most are damn hard, but this one should be a piece-of-cake."*
Strategy 4: Use ellipsis (omission) to communicate tacit understandings	*"(Do you) mind if I join you?"*
Strategy 5: Use first name or in-group name to insinuate familiarity	*"Hey Bud, have you gotta minute?"* *"Hey Johnny, how's it going?"*
Strategy 6: Claim common view; assert knowledge of hearer's wants	*"You know how the janitors don't like it when..."*
Strategy 7: Seek agreement; raise or presuppose common ground/common values	*"I'll be seeing you then."* *"So when are you coming to see us?"*
Strategy 8: Hedge opinions; be vague about disagreeing opinions	*"It's really beautiful, in a way."* *"I don't know, like I think people have a right to their own opinions."*
Strategy 9: Engage in small talk/joke	*"How bout that game last night? Did the Ravens whip the pants off the Giants or what!"*
Strategy 10: Give or ask for reasons; make activity seem reasonable to the hearer	*"I'm really late for an important appointment, so ..."* *"Why not lend me your cottage for the weekend?"*
Strategy 11: Use inclusive forms to include both speaker and hearer in the activity	*"We're not feeling well, are we?"* *"I really had a hard time learning to drive, you know."* *"Let's get on with dinner"*
Strategy 12: Use presuppose manipulation by presupposing wants, attitudes, values	*"Wouldn't you want a drink?"* *"Look, you're a pal of mine, so how about…"*
Strategy 13: Assert reciprocal exchange or tit for tat	*"Do me this favor, and I'll make it up to you."*
Strategy 14: Be optimistic and presume cooperation	*"Look, I'm sure you won't mind if I borrow your typewriter."* *"I just dropped by for a minute to invite you for tea tomorrow - you will come, won't you?"*
Strategy 15: Give something desired - gifts, sympathy, understanding	*"You look like you've had a rough week."*

Negative Politeness Strategies	
Strategy and Description	**Examples**
Strategy 1: Be conventionally indirect; inquire into the hearer's ability or willingness to comply	*"Can you tell me what time it is?"* *"Could you possibly pass the salt, please?"* *"Are you by any chance able to post this letter for me?"* *"I hope that you'll please close the door."* *"There wouldn't, I suppose, be any chance of your being able to lend me your car for just a few minutes, would there?"*
Strategy 2: Use hedges: words or phrases that diminish the force of a speech act	*"Can I perhaps/possibly trouble you?"* *"I suppose/guess/think that Harry is coming."* *"Would you close the window, if you don't mind?"* *"This may not be relevant/appropriate, but…"* *"Since I've been wondering…"* *"By the way…"*
Strategy 3: Use the subjunctive to express pessimism about the hearer's ability/willingness to comply	*"Could (instead of can) I ask you a question?"* *"I don't imagine/suppose there'd be any chance/possibility/hope of you…"* *"Perhaps you'd care to help me."*
Strategy 4: Use words or phrases that minimize the imposition, like "just" or "a little"	*"I need just a little of your time."*
Strategy 5: Give deference by using honorifics: Sir, Mr., Ms., Dr.	*"Can I help you, Sir?"*
Strategy 6: Use formal word choices to indicate seriousness and to establish social distance	*"I look forward very much to dining (instead of eating) with you."*
Strategy 7: Apologize: admit the impingement, express reluctance	*"I am sorry to bother you, but…"* *"I'm sure you must be very busy, but…"* *"I'd like to ask you a big favor…"* *"I hope this isn't going to bother you too much…"* *"I know you've never bothered me, but…"*
Strategy 8: Give overwhelming reasons	*"Can you possibly help me with this, because there's no one else I could ask."* *"I simply can't manage to…"*
Strategy 9: Impersonalize the speaker and hearer by avoiding the pronouns "I" and "you"	*"Is it possible to request a favor?"* *"It appears (to me) that…"* *"It would be appreciated if…"* *"We regret to inform you…"*
Strategy 10: Use the past tense to create distance in time	*"I had been wondering if I could ask a favor."* *"I hoped I might ask you…"*
Strategy 11: Nominalize (change verbs and adverbs into adjectives or nouns) to diminish the speaker's active participation	*"My asking you to leave is required by regulations."* *"It is my pleasure to be able to inform you…"*
Strategy 12: State the FTA (Face-Threatening Act) as a general rule	*"Regulations require that I ask you to leave."*

Negative Politeness Strategies (continued)	
Strategy and Description	**Examples**
Strategy 13: Go on record as incurring a debt	*"I'd be eternally grateful if you would…"*

B Scenarios

Scenarios highlighted in bold are equivalent to the ones used in the writing scenarios

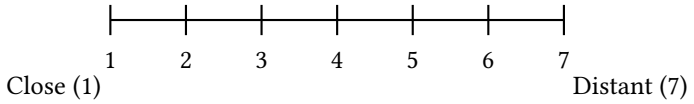
Social Scenarios		
Number	**Type**	**Scenario**
1	**Middle**	**You're working on a course project for a class and attended a TA-led review session last week. You send an email asking the TA for clarification on how to apply a key concept discussed during the review to your project, and if they could point you toward any relevant examples or references that might help.**
2	**Middle**	**You and a colleague are preparing for an upcoming team presentation. You've completed your part, but your colleague is responsible for compiling all the sections into a final document. You send them a message to see if they need any support finishing up the compilation and to check whether they'll be able to share the final draft before the weekend so there's time for review.**
3	**Middle**	**You're part of a research group for a university course, and you've been meeting regularly with your project advisor. You write an email to request feedback on your recent data analysis and ask if they could recommend any tools or references to enhance your final report.**
4	Middle	You're working on a project that involves coordinating with other teams. You ask a colleague from a different department, whom you've worked with occasionally and whose role includes handling interdepartmental communications, if they could help review and ensure that all necessary information is properly prepared before you finalize your report for submission to your supervisor.
5	Middle	You need to request a meeting with a team member in your department to discuss potential collaboration on a project. You've interacted a few times but haven't worked closely together.
6	Middle	You are starting a research project and you recall a senior researcher in your department, with whom you've had several conversations during department events, whose expertise closely aligns with your topic. You believe they might be interested in collaborating but are uncertain about their availability.
7	Low	You are going to a conference next week, and you know one of your colleagues are going to the same conference and they also don't have a car. You write an email to ask if it's possible to share a ride to the venue.
8	Low	You've recently taken up a new hobby, hiking, and you know that your colleague enjoys similar activities. You're writing an email to ask if he/she'd like to join you for a hike this weekend.
9	High	You ask your professor who taught a large lecture class you attended a few years ago but with whom you had little direct interaction, to introduce you to someone who may be hiring in your chosen career path.

Social Scenarios (continued)		
Number	**Type**	**Scenario**
10	High	You are organizing a major event and need a high-profile sponsor to secure the necessary funding. You decide to reach out to a CEO of a leading company, whom you met briefly at a fundraising event, to request sponsorship.

C Distance, Power and Imposition Measurements

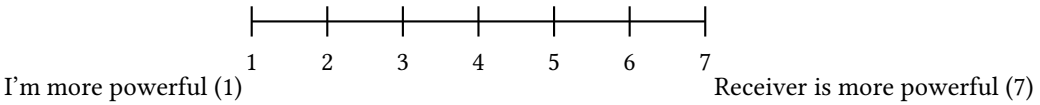
C.1 Distance

How socially distant do you feel with the receiver of the email?
Social distance refers to how close or familiar two people feel with each other. It's often based on how often they interact and what they exchange, like help, favours, or emotional support.



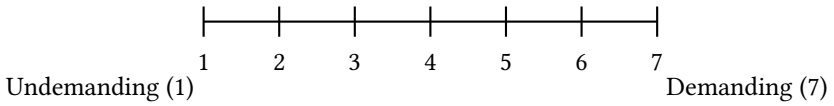
C.2 Power

How do you perceive the power dynamic between you and the receiver?
Power dynamic refers to how much one person can control or influence another. This can involve control over resources (like money or authority) or the ability to shape others' beliefs and actions.



C.3 Imposition

How demanding do you think your request is to the receiver?
The level that a task is **demanding** can be determined based on two factors: 1) the services requested (like time or effort) and 2) the goods requested (including non-material goods like information or expressions of respect).



D Email Quality Rubric

Email quality refers to effective written communication achieved through:

- Clear language and grammatical accuracy.
- Appropriately structured email format.
- Appropriate use of communication strategies to which the reader would respond positively.

Below are example characteristics of high-quality versus low-quality emails. Judgments should be based on overall impressions, assuming the role of the reader.

1 - Extremely Low Quality:

1. The language is inaccurate, with major grammatical errors.
2. The email is inappropriately structured, such as:

- Missing salutation or closing.
- Salutation or closing is not context-appropriate.

3. The email lacks sufficient information for the reader to understand the context of the request, for example:
   - Missing self-introduction.
   - Lacking a detailed description of the matter of concern.
4. The email is overly wordy and repetitive, making it difficult for the reader to extract relevant information.

Overall, the email is hard to read, indicating the writer's lack of ability or effort to communicate their request effectively.

7 - Extremely High Quality:

1. The language used is accurate with minimal grammatical errors.
2. The email is appropriately structured, including context-appropriate salutations and closings.
3. All necessary information is included for the reader to understand the context of the request, such as:
   - Providing self-introduction with context-relevant details if the scenario requires.
   - Describing the matter of concern with necessary details.

Overall, the email is easy to read, makes the reader feel respected, and conveys that the writer is capable and has put in sufficient effort to draft a high-quality email. The reader feels positive about the email writer.

E Methodological Note on Experimental Design

In this note, we clarify the study design decisions and address differences in methodological expectations across disciplines. In this study, we adopted a within-subject design for the independent variable AI POLITENESS, comparing two conditions, *positive-AI* and *negative-AI*. The within-subjects design (also referred to as repeated-measures design) is commonly used in HCI and psychology research [36] [52] [51]. It offers the benefit of requiring a smaller subject pool and less individual variance. However, we acknowledge that interpretations of experimental rigour may differ across disciplines and therefore include this note to bridge these differing methodological perspectives.

Our study aims to establish causal relationships between different AI writing assistant configurations and their effects on writing strategies and social perception. The purpose of a controlled experimental design is to create conditions under which observed results can reasonably support causal interpretation. Establishing causal inference typically requires three conditions: association, temporal order, and consideration of alternative explanations [2, 50]. In this study, association is examined through the relationship between the manipulated independent variable and the dependent variables. Temporal order is established by the study procedure, in which users interact with a specific AI writing assistant configuration before producing email responses and evaluations. As a result, observed differences occur after exposure to the experimental conditions.

Addressing alternative explanations is a central concern in experimental design, and within-subject and between-subject designs each offer distinct strengths and limitations. A key alternative explanation in our context is variation in participants' baseline tone and writing style. The within-subject design directly addresses this concern by having each participant complete all conditions, which supports comparisons within individuals rather than across different participants. On the other hand, between-subject designs are less effective at controlling for individual variability because each condition includes different participants. For this reason, conventional between-subject experiments rely on random assignment to conditions to reduce systematic confounds.

Using a within-subject design can also introduce additional confounds, such as order effects, because the same participant is exposed to multiple conditions. Participants may perform better in later conditions due to increased familiarity with the task. A between-subjects design can reduce this concern by assigning each participant to only one condition. To mitigate order effects in our within-subject design, we randomized the order in which participants experienced the *positive-AI* and *negative-AI* conditions. After weighing these trade-offs, we chose a within-subject design because we expect individual differences to play a substantial role in our study, while potential order effects are less likely to meaningfully influence our outcomes.

Finally, this study compares two AI writing assistant configurations rather than contrasting AI assistance with a non-AI condition. This choice follows directly from the research question, which focuses on how different implementations of politeness strategies in AI-mediated writing influence user behaviour and perception. Comparative condition designs of this kind are common in HCI and system evaluation research and are well-suited for isolating the effects of specific design choices within interactive systems.

Received May 13, 2025; revised January 13, 2026; accepted April 9, 2026