Int. J. Human–Computer Studies 210 (2026) 103770

Contents lists available at ScienceDirect

International Journal of Human - Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

Culture, emotions, and power dynamics in AI email communication

Marina Polupanova [a], Marios Constantinides [b, c, *], Daniele Quercia [b]

[a] Nokia Bell Labs Consulting, Bellevue, Washington, United States
[b] Nokia Bell Labs, Cambridge, United Kingdom
[c] CYENS Centre of Excellence, Nicosia, Cyprus

HIGHLIGHTS

- Large Language Models (LLMs) enhance email clarity but do not necessarily increase recipients' willingness to act.
- Willingness to act improves when email requests are perceived as trustworthy and appreciative.
- Effective emails require positive elements and avoidance of potentially toxic statements.

ARTICLE INFO

Keywords:
Large language models
Email communication
Emotions
Culture

ABSTRACT

Large Language Models (LLMs) have shown remarkable capabilities in enhancing writing tasks. However, the effectiveness of LLMs in specifically improving email communication is yet unclear not least because such communication depends on complex factors, including power dynamics, cultural contexts, and emotional nuances. To understand the effectiveness of writing emails with LLMs, we conducted a crowd-sourcing study with 266 participants who annotated two sets of emails in terms of clarity and willingness to act upon the task communicated in those emails: one set was generated by humans and another generated by LLM. The emails were based on four use cases, including requests for data gathering and analysis, workload estimation, and meetings or events organization, representing situations with the same or different power levels between the sender and receiver, and were developed using best practices for project management. We found that, on average, LLM-generated emails expressed tasks in a clearer way but were not more likely to result in willingness to act. In fact, willingness to act increased not only if the receiver is a subordinate but also if the receiver perceives the request to be trustworthy and appreciative. Our crowdsourcing approach allowed us to understand the effectiveness of AI-generated text for email communication, and our findings suggest that future AI-assisted innovations should prioritize fostering trust, appreciation, and respect between communicators—factors that will remain decisive for cooperation even if AI systems become more advanced or approach superhuman capabilities.

1. Introduction

Effective email communication has always been crucial in the workplace. McKinsey research suggests that spending 7–8 min to enhance an email's clarity and grammar can increase productivity by up to 30% (McKinsey, 2012). Clearly articulating tasks in emails ensures that receivers understand and can act on them without confusion, whereas ambiguities can lead to prolonged clarification exchanges or incorrect actions (Sappelli et al., 2016; Lampert et al., 2008; Cohen et al., 2004). This principle also extends beyond email communication: unclear task instructions in crowdsourcing studies have necessitated the development of systems that help workers understand and execute tasks effectively, thus maintaining high-quality results with minimal dependence on the clarity provided by requesters (Manam and Quinn, 2018).

Recent advances in generative AI have introduced new dynamics into workplace email communication. Large Language Models (LLMs), such as GPT-4, are increasingly employed to compose or refine professional messages. These tools are often acknowledged for enhancing fluency and clarity, particularly for non-native speakers or writers inexperienced early-career professionals (Cambon et al., 2023; Dell'Acqua et al., 2023; Dhillon et al., 2024). However, their use also raises concerns about authenticity, trust, and emotional nuance. Several studies have found that recipients may perceive LLM-generated messages as insincere or inauthentic (Hohenstein et al., 2023a; Hoque et al., 2024), while

* Corresponding author at: Nokia Bell Labs, Cambridge, United Kingdom.
Email address: marios.constantinides@nokia-bell-labs.com (M. Constantinides).

https://doi.org/10.1016/j.ijhcs.2026.103770
Received 27 August 2024; Received in revised form 28 August 2025; Accepted 16 February 2026
Available online 19 February 2026

others report that individuals often struggle to distinguish AI-generated content from human-written text and rely on flawed heuristics to make such judgments (Jakesch et al., 2023). Moreover, Kadoma et al. (2023) observed that although the style of the AI-generated message did not significantly influence perceptions of inclusion, participants who felt more included also reported greater agency and ownership, especially among members of minority groups.

The use of generative AI in communication is not purely technical; it alters the nature of authorship, intent, and interpersonal alignment. Recently, Constantinides et al. (2025) introduced the concept of "blended work" to highlight how human-AI collaboration reshapes users' sense of control, authenticity, and identity. As generative AI becomes embedded in daily communication workflows its impact on interpersonal relationships and how users prompt, trust, or reinterpret it demands closer empirical study.

Beyond writing quality, effective email communication depends heavily on interpersonal dynamics (e.g., power relationships, cultural expectations, and emotional tone). Emails written by a supervisor or by a more senior or a competent employee may be interpreted differently than those written by an equal-rank peer (DeWall et al., 2011). Likewise, email styles vary across cultures in terms of formality, directness, and emotional expressiveness (Holtbrügge et al., 2013; Meyer, 2014). Additionally, the way emotions are expressed and interpreted in emails can lead to misunderstandings; (Byron, 2008) found that employees often perceive emails' emotional content as more negative or neutral than intended.

While prior research has examined the linguistic clarity and perceived fluency of LLM-generated messages, much less is known about whether such messages motivate recipients to take action to the same extent as human-written ones, how subtle differences in emotional framing influence recipient motivation, and whether prompting strategies can help AI writers convey interpersonal intent more effectively. To address these gaps, we make two main contributions:

1. We conducted a cross-cultural crowd-sourced study (n = 266) comparing human- and LLM-generated emails, assessing how message source, perceived emotions, cultural alignment, and power dynamics influence recipients' ratings of clarity and willingness to act (Section 4).
2. We found that while LLM-generated emails were clearer than human-written ones, clarity alone did not predict willingness to act (Section 5). Instead, perceived emotional tone was the strongest predictor, with power dynamics playing a secondary role and cultural alignment showing no effect. Additionally, through thematic analysis, we identified phrasing strategies that mapped to specific emotional interpretations (e.g., trust *vs.* disgust) and influenced willingness to act.

Our findings contribute to the emerging literature on human–AI communication and culturally-aware AI design (Section 6). They demonstrate which prompting strategies can guide LLMs to produce text that elicits desired emotional responses and fosters recipient cooperation in email communication.

2. Related work

Despite the growing adoption of LLMs in the workplace, understanding how AI-generated messages are interpreted across cultures and organizational hierarchies remains fragmented. Our study builds on three intersecting domains: *emotional perception in mediated communication*, *cultural and power asymmetries in workplace discourse*, and *LLM prompt engineering*. Together, these fields suggest rich possibilities for AI-enhanced messaging but also highlight unresolved tensions that our study addresses.

Emotional Perception and Action in Mediated Communication. Emotions, expressed or inferred in text-based communication have long been recognized as critical to recipient responses. Early studies in computer-mediated communication (CMC) showed that emotional cues, even when implicit, shape compliance, cooperation, and trust Hancock et al. (2007); Byron (2008). Politeness markers, intensifiers, hedging, and acknowledgments serve as affective signals that influence whether a message is perceived as respectful, urgent, or neglectful.

Contemporary affective lexicons (e.g., Mohammad and Turney, 2013; Mohammad, 2018) systematize these cues, while dimensional emotion models such as those proposed by Plutchik (1982, 2004); Gunes and Pantic (2010) offer taxonomies for classifying textual affect. Linguistic variability has been explored in learning contexts by Kort et al. (2001)), who identified scales such as anxiety-confidence and frustration-euphoria, and discussed strategies to transition from one emotional state to another in the learning process of memory recall. Recently, Jackson et al. (2019) challenged universal emotional categories, and suggested that emotion words may carry different meanings across cultures and languages. Boroditsky (2011) established that language shapes perception, and illustrated this in how native English, Japanese, and Spanish speakers remember events. Semin et al. (2002) discovered that representatives of interdependent cultures (e.g., Hindustani, Turks) express emotions mostly with verb-based relational phrases, while independent cultures (e.g., Dutch) do so with noun or adjective-based abstract phrases. Argaman (2010) shows that strong emotional expression linguistic markers include intensifiers, emotional vocabulary, repetitions, self-references and words derived from "feel", and their use increases with emotional intensity. At the same time, emotionally strong messages contain fewer unique words than less emotionally intensive.

These findings suggest that perceived emotion is not simply a property of a message, but a dynamic interaction between wording, culture, and context. The emotional component of messages should not be studied in isolation, as Russell (1991) noticed that emotion interpretation may have a culture-specific component. In addition to that, classic theories of power distance (Hofstede, 2011) and high- *vs.* low-context communication (Meyer, 2014; Hall, 1976) reveal that authority and indirectness are not interpreted universally. This led us to include emotion and culture variables together with thematic analysis of workplace originated texts in our study. Moreover, relatively few studies link emotion perception directly with willingness to act, especially in AI-generated text, where emotional resonance may differ even when clarity is preserved. Similarly, most emotion detection efforts remain focused on sender intent, rather than receiver interpretation; a distinction that becomes crucial in LLM-mediated contexts, where authorship may be shared, obscured, or fully synthetic.

Cultural and Power Asymmetries. The emotional component of messages should not be studied in isolation, as Russell (1991) noted that emotion interpretation may have a culture-specific component. In addition to that, classic theories of power distance (Hofstede, 2011) and high- *vs.* low-context communication (Meyer, 2014; Hall, 1976) reveal that authority and indirectness are not interpreted universally. For example, in high power distance cultures (e.g., India, Japan, Poland), workers often defer to positional authority, while low power distance cultures (e.g., the US, UK, Germany) favor flatter hierarchies and more participatory communication styles (Khatri, 2009; Ghosh, 2011). Another example would be that participants from Germany collaborated more equally, while those from Poland emphasized hierarchy in collaborative innovation projects (Avelino et al., 2023). However, even within hierarchical structures, psychological ownership can override power distance effects and allow employees to have more agency in goal setting (Pervaiz et al., 2024).

In multicultural teams, misalignment in tone or structure can lead to unintended emotional responses that range from confusion to offense (Holtbrügge et al., 2013; Avelino et al., 2023). For example,

Meyer (2014) noted that Americans often misinterpret indirect or nuanced communication styles common in East Asian contexts. Dong et al. (2016); Wiggins (2012) also explored how intercultural relationship modeling can mitigate such challenges.

These dynamics persist in email even when the content is functionally clear. Prior research has shown that the perceived status of the sender can moderate how a message is received (DeWall et al., 2011), and that culturally incongruent styles can undermine cooperation. However, current LLMs, trained on heterogeneous data, often default to dominant (e.g. Western, low-context) communication styles (Farid Adilazuarda et al., 2024). While some work aims to improve LLMs with cultural competence (Li et al., 2024), most empirical studies still assess LLM output in isolation, rather than in interactive contexts where power and culture are inferred by the recipient. Our study addresses this gap by treating both power cues and cultural alignment as perceptual phenomena, shaped by emotional tone and phrasing, and analyzing how these shape behavioral intentions.

LLM prompt engineering. Prior work on generative AI in email communication has mostly focused on efficiency and clarity gains (Kannan et al., 2016; Das et al., 2021; Dell'Acqua et al., 2023). Experiments with GitHub Copilot (a tool that turns natural language prompts into coding suggestions) (Peng et al., 2023) and task-focused automation (Kokkalis et al., 2013) support the hypothesis that AI augments productivity. Yet concerns persist. Hoque et al. (2024); Hohenstein et al. (2023b) found that AI-generated emails are often viewed negatively due to concerns about authenticity. Moreover, concerns remain about cultural biases in AI-generated text (Kolisko and Anderson, 2023; Venkit et al., 2023), and the data used for its training (Inel et al., 2023; Hsu et al., 2022) indicating the need for careful implementation and monitoring (Bringas Colmenarejo et al., 2022). Previous research showed that workers involved in AI training perceive these concerns seriously, and are even ready to incur some costs themselves for improving AI fairness (Treiman et al., 2023). Additionally, there is evidence that humans tend to perceive the quality of AI-generated decisions negatively (Grgić-Hlača et al., 2022).

Nevertheless, it has been shown that (generative) AI can significantly improve and augment human productivity (Dell'Acqua et al., 2023). In an experiment using GitHub Copilot, software developers completed tasks 55.8% faster with the assistance of an AI pair programmer compared to a control group without AI assistance (Peng et al., 2023). Kokkalis et al. (2013) demonstrated how extracting the tasks from emails with AI tool for crowd-sourcing workers allowed them to increase their task completion rate almost twofold (from 29.3% to 58.4%). Another study using deep-learning algorithms on Enron[1] email exchanges revealed optimal working hours and productivity patterns (Das et al., 2021). Similarly, one may hypothesize that writing emails with AI can enhance overall productivity, leading to effective email communication (Nugroho et al., 2023). Jovic and Mnasri (2024) studied the efficiency of various chatbots (ChatGPT-3.5, Bard, Llama 2.0, and Bing Chat) in conveying messages to recipients based on customer experience scenarios (e.g., refusal of a refund and client complaints) and typical office situations (e.g., a request to the manager to approve remote work). The most consistent results were observed in the Llama model; however, it was reported that all models struggled with understanding emotional context and technical aspects of the messages. The research made conclusions on the efficiency using a specifically designed rubric and the grading was done by the researchers. Apart from enhancing the clarity of messages, researchers have also studied the ability of AI to convey calls to action. Carrasco-Farre (2024) noticed that lexical and grammatical complexity of the AI-written texts, along with the moral language used by a particular LLM, could be potential reasons why AI-generated content can be as persuasive as content written by humans.

[1] Enron was an American energy, commodities, and services company.

Prompt engineering has emerged as a critical tool for steering LLM output, not just in content but in tone, emotion, and relational positioning. Studies show that even minimal changes in prompts can induce substantial shifts in affect and coherence Dhillon et al. (2024); Salinas and Morstatter (2024). As these findings often rely on expert ratings or benchmark tasks rather than on the writer's intended emotions or stance, it is important to note that Buechel and Hahn (2022) demonstrated the dominance of the reader's emotional perspective in achieving high inter-annotator agreement among experts. Similarly, Braun et al. (2020) found that readers tend to perceive both negative and positive emotions in written text more strongly than the author intended, and that agreement is generally higher among annotators themselves than between authors and annotators.

Gandhi and Gandhi (2025) argue that prompt sentiment serves as a "catalyst" for stylistic control, yet they stop short of identifying which linguistic forms correspond to specific emotional outcomes in naturalistic writing. Additionally, although sampling parameters such as temperature influence the variability of LLM outputs Renze (2024), most non-expert users do not adjust them. Instead, they rely on prompt rewriting, as public LLM tools rarely provide easy-to-use controls for changing model parameters.

Our linguistic analysis contributes to this emerging literature by identifying which phrasing strategies (in both human- and LLM-generated messages) correlate with specific perceived emotions and higher willingness to act. Through thematic analysis, we identified how gratitude, criticism, support, and time-respect cues shape perceived emotions such as trust, disgust, or fear.

Research Gaps. In summary, previous works have mainly studied the clarity and usefulness of AI-generated messages(Peng et al., 2023), the cultural variation in email interpretation (Meyer, 2014; Holtbrügge et al., 2013), the influence of emotional tone on behavior (Brundin et al., 2008; Hancock et al., 2007), and the role of prompting in AI output tone (Dhillon et al., 2024). However, only a few studies have identified how LLMs impact motivation of readers to perform a task, particularly when their emotional tone is misaligned or socially ambiguous, and which linguistic strategies (especially in AI-mediated text) map to specific emotional interpretations and behavioral responses in real-world tasks.

To address these gaps, we conducted a cross-cultural, crowd-sourced study comparing human and LLM-generated emails. Our work analyzes not only recipient ratings of clarity and willingness to act, but also the underlying emotional, cultural, and hierarchical cues that shape those outcomes. Through thematic analysis, we identified concrete phrasing patterns that correspond to specific emotional reactions, and offer practical insights for prompt designers and AI writing support tools.

3. Author positionality statement

We situate ourselves in a Western country during the 21st century, writing as authors primarily engaged in academic and industry research. Our team comprises one female and two males from Southern and Eastern Europe with diverse ethnic and religious backgrounds. Our combined expertise covers a range of areas, including human-computer interaction (HCI), ubiquitous computing, software engineering, artificial intelligence, project management, and telecommunications. As researchers from a primarily Western institution, we recognize the need to expand the methodological perspectives discussed in this paper and take responsibility for the content and interpretations of our findings. Our diverse backgrounds and institutional context likely shaped our approach to selecting methodology, designing our study, and interpreting the data.

4. Methodology

Prior studies have demonstrated productivity gains without accuracy loss for regular office tasks when office workers are augmented with AI (Dell'Acqua et al., 2023; Cambon et al., 2023). However, the extent

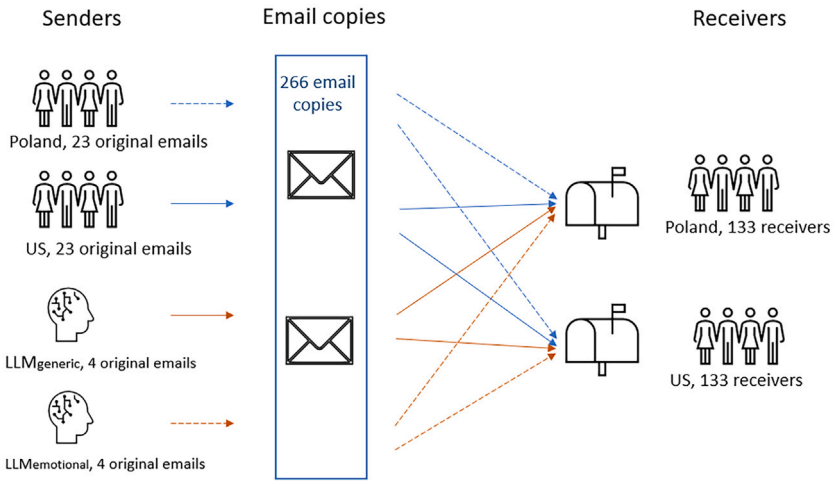


Fig. 1. Send-and-receive email procedure. Unique emails generated by Polish (23 emails), US (23 emails) and LLM senders (8 emails). These emails were replicated and sent in a random order to 133 receivers from Poland and 133 receivers from US.

to which AI can help in email communication and the role of power dynamics, culture and emotions in this process has not been fully studied. To address this gap, we set out to study whether power dynamics, cultural and emotional factors are associated with effective email communication, and the extent to which generative AI technology can help improve it. To this end, we formulated 2 Research Questions (RQs):

RQ_1: Do human-generated emails differ from LLM-generated ones in terms of task clarity and willingness to act upon an email?
RQ_2: What is the role of cultural background, power differences, and perceived emotions in people's willingness to act upon emails, and to what extent do LLMs' prompting strategies relate to the impact of these factors?

To answer our RQs, we conducted a crowd-sourcing study wherein participants (called "receivers") annotated two sets of emails (one generated by humans—called "senders", and another generated by LLM) in terms of clarity and willingness to perform the task communicated in those emails. The study was approved by Nokia Bell Labs. Next, we describe the crowd-sourcing study setup.

4.1. Participants

We used Prolific[2] to recruit participants, considering a number of criteria including cultural differences, work expertise, and level of education.

To account for cultural differences, we recruited participants from two "culturally distant" countries, that is, the United States (US) and Poland. This choice was based on two factors. First, according to Meyer (2014), these two countries have different cultures, which can significantly influence the way individuals perceive and interact with the world. For example, Poles are considered to be more high-context (i.e., relying more on the non-verbal cues and implicit messaging), relationship-based and direct in giving feedback than individuals from the US (Meyer, 2014). Hofstede (2011) mentions that Poland, with its Central European heritage, often exhibits more collectivist tendencies where community and family ties play a more prominent role in decision-making and social behavior. In contrast, in Hofstede's terms, the US is typically characterized by individualistic values, emphasizing personal achievement and independence (Hofstede, 2011). Also, according to (Meyer, 2014), the US demonstrates low-context communication, where spoken or written information explicitly contains the meaning of the message.

Second, after examining several "culturally distant" country pairs as per (Meyer, 2014)—including Mexico-US, India-US, and France-US—we found that, at the time of executing the study, only the Poland-US pair had a sufficient number of Prolific participants to achieve the statistical power necessary for our analysis. Additionally, we required that participants spend most of their time until the age of 18 in their respective country, thus controlling for any language bias. It is important to state that prior studies have shown that humans acquire directional bias only after becoming literate in their native language during school days (Dobel et al., 2007). This bias manifests in all languages so that the spatial positioning of an English native speaker and any other language native speaker expressed in English language will differ. Since schooling can be completed by the age of 18 in most countries, we assume that if an individual remained in their home country until that time, they would demonstrate positional or other biases, which are influenced by their native language when writing in English.

For work expertise, we chose knowledge workers by limiting the industries where participants work to information technology, telecommunication, and finance, and the level of education equal to or above college level. Both work expertise and level of education were used as controls for participants who are more likely to engage with corporate email communication. However, no restriction was implied on the participants' position in corporate hierarchy or current employment status.

Participants' ages ranged between 18 and 65 years, with 31 years as the median age for both senders and receivers. Due to the randomized email send-and-receive procedure (Fig. 1) and data quality checks, we did not aim to achieve absolute gender parity among the respondents. Instead, we aimed for minimum possible difference between gender counts. In total, our sample included 17 female and 29 male senders, and 125 female and 141 male receivers. Each participant was compensated at an average hourly rate of $11, and paid in their local currency. This rate was higher than the minimum rate of $8 recommended by Prolific for both countries. The average completion time for senders' tasks was 12.6 min, and for receivers' tasks was 9.27 min. Demographics of the senders and receivers are provided in Tables 1 and 2.

4.2. Materials

To prompt participants to write emails, a set of 4 use cases was constructed by a project management professional (certified practitioner) with more than 15 years of experience in a large technology company,

[2] https://www.prolific.com/.

Table 1
Sender's country of residence, gender, median age, education, and the number of written emails.

Country	Gender	Median age	Education	Email count
Poland	Female	25	Graduate degree	3
			Undergraduate degree	4
			Technical/community college	2
	Male	27	Doctorate degree	1
			Graduate degree	2
			Undergraduate degree	7
			Technical/community college	4
US	Female	36	Graduate degree	2
			Undergraduate degree	5
			Technical/community college	1
	Male	32	Graduate degree	6
			Undergraduate degree	8
			Technical/community college	1

Table 2
Receivers' country of residence, gender, median age, education, and the number of emails they rated.

Country	Gender	Median age	Education	Email count
Poland	Female	29	Graduate degree	35
			Undergraduate degree	20
			Technical/community college	5
	Male	26	Doctorate degree	1
			Graduate degree	31
			Undergraduate degree	20
			Technical/community college	21
US	Female	37	Doctorate degree	4
			Graduate degree	7
			Undergraduate degree	38
			Technical/community college	16
	Male	36	Doctorate degree	2
			Graduate degree	11
			Undergraduate degree	41
			Technical/community college	14

and further verified by two experienced project managers from our institution. According to the Project Management Body of Knowledge book (PMBOK 7th edition), the four use cases can be mapped to the following 'methods' domains (Project Management Institute, 2021): Data Gathering and Analysis (Case 1 and Case 2), Estimating (Case 3), Meetings and Events (Case 4).

Cases were designed in a way that allowed testing the differences in clarity, willingness to act, and perceived emotions not only for two culturally diverse groups, but also for situations where the sender and receiver have same or different power levels. Each case contained the following:

- Task description (e.g., provide a spreadsheet with costs, a Word document with system details)
- Sender's position and qualifications, if any (e.g., experienced Solution Architect)
- Receiver's position and qualifications, if any (e.g., Project Manager)
- Potential emotional trigger for the sender (e.g., past performance of the receiver on a given task).

Out of four cases, one (Case 1) had the same power level for sender and receiver. The remaining three cases (Cases 2-4) displayed a higher power difference between sender (manager) and receiver (subordinate). Table 3 describes the main features of the four use cases, and their detailed descriptions are available in the Appendix.

4.3. Procedure

The study included two steps with human involvement: writing emails by a group of "senders" (generating emails[3]) and processing them by another group of "receivers" (responding to emails). Note that we did not allow a receiver to be a sender. Also, the receiver was not aware of cultural, educational, or any other background of sender and was not aware whether the email was written by human or AI. The procedure is shown in Fig. 1 and the interface in which both senders and receivers performed the requested tasks is shown in Fig. 2.

Generating emails (senders). For human-generated emails, each sender was presented with a description of a randomly selected use case, and was tasked with writing an email of 150–250 words in length to the person mentioned in the case. We explicitly instructed participants not to use generative AI tools (e.g., ChatGPT) for writing, to spend no more than 5–9 min on the task, and to use all relevant context from the use case's description. We did not prescribe emotional tone, persuasion strategies, or linguistic style, as our aim was to capture the natural variance in affective and interpersonal expression. Subsequently, emotions and tone were evaluated through third-party perception (subsequently described in "Responding to emails"). We did not attempt to ask senders what emotions they wanted to express in the email, and captured only perceived rather than intended message affect, as Braun et al. (2020) shows that emotional perception of the message by senders and readers can differ and for us perceived emotion is more important as it may impact the receiver's motivation to perform the task.

For LLM-generated emails, we used the same use cases as input to OpenAI's GPT-4 model. We chose GPT-4 because it was the best performing model[4] (OpenAI, 2023) at the time of executing the study in November–December 2023. We used two prompting strategies:

- Prompted simply with *"write a 150–250 word email using the information in the use case"*. We labeled emails generated like that as $LLM_{generic}$
- Prompted with *"write a 150–250 word email using the information in the use case, and include explicit descriptions of emotions regarding the situation described in the use case"*. This portion of emails was labeled as $LLM_{emotion-aware}$

As recent research has shown, small changes in prompt design can significantly impact semantic output and tone (Salinas and Morstatter, 2024), and users' prompting strategies act as a crucial mediating layer in AI-augmented communication. As an LLM can already amplify a prompt's emotional bias (Gandhi and Gandhi, 2025) caused by emotionally-provoking situations in our use cases, we chose two prompting strategies above to define whether a regular prompt with no reference to any emotions, or an explicit request to express some emotions will be more efficient for motivating a reader to complete the task.

We intentionally generated just one email per use case for each of the two prompts and used default Chat GPT temperature and top_p parameter values to simulate a typical user experience with publicly available LLM tools where changing those parameters is not a typical user behavior. Replicating precise generation conditions months later may become infeasible, making our focus on prompt design a more stable and generalizable point of comparison. Also, Renze (2024) shows limited effect of temperature changes on LLM prompting results, while

[3] In addition to human-generated emails, we produced another set of LLM-generated emails.
[4] https://lmsys.org/blog/2023-06-22-leaderboard/.

Table 3
The four use cases' characteristics; power dynamics (i.e., sender and receiver positions), the potential emotional trigger for the sender, and the task description.

Use case	Sender position	Receiver position	Emotional trigger	Task description
1	Solution Architect	Solution Architect	Sender knows that the receiver's project in the past was done successfully.	To share receiver's past experience in the similar project (risks, limitations, extra integration steps)
2	Project Manager	Project Team member	PM suspects that there will be cost overrun in the project	To download workload and financial raw data from corporate tool and make a summary table
3	Consultant Engagement Lead	Domain Experts	One of the experts failed the similar task last time and the team lost the bonus	To estimate the workload and describe the list of tools for the new project based on the past experience
4	Program Manager	Project Manager	Program manager heard a good feedback on Project manager's performance for past projects	To give the Project manager reporting guidelines and encourage him to use best practices of Program manager team for customer calls preparation.

Fig. 2. Qualtrics survey used in the study. (a) Example scenario description shown to senders, who were asked to compose an email in response. (b) Interface shown to receivers to rate the clarity of a randomly assigned email. (c) Follow-up interface where receivers indicated their willingness to perform the described task. (d) interface where receivers labeled the emotions they perceived in the email, choosing from a predefined list or entering their own. All tasks were presented sequentially in Qualtrics, with receivers viewing only one email each.

Salinas and Morstatter (2024) opts to use temperature = 0. Examples of emails generated both by humans and LLM are provided in Appendix A.3.

Responding to emails (receivers). After obtaining both the human- and LLM-generated emails, we randomly assigned them to a country-balanced group of receivers. Each receiver had seen only one email from an entire email corpus (to prevent carry-over effects or contrast bias), and was tasked to:

- Rate the email's clarity on a Likert scale from 1–10 (1 indicating "vague", 5 indicating "neutral", and 10 "clear").
- Rate the email's willingness to perform the task expressed in the email on a Likert scale from 1–10 (1 indicating "not willing", 5 indicating "neutral", and 10 "willing"). These scores were scaled between 0 and 100.
- Pick from a pre-defined list of eight emotions all emotions which they believe the sender is experiencing towards them. In case there was no label in the list which best expressed the sender's attitude towards the receiver, they had the opportunity to propose new one.

Examples of how those questions to a receiver looked in the poll UI are displayed in Fig. 2. This third-party annotation approach reflects best practices in emotion perception research, especially where sender intent is unknown or not explicitly encoded, such as in short messages or mediated communication (Mohammad, 2018; Gunes and Pantic, 2010.

To ensure statistical power in our analyses, we first determined the number of receivers needed. Using the F-test ANOVA from the G*Power tool (Buchner, Axel and Erdfelder, Edgar and Faul, Franz and Lang, Albert-Georg , 2007), we found that with an expected effect size of 0.3, a sample size of around 256 receivers would be sufficient to achieve statistical power of 95% at a significance level of 0.05. Then, we determined

Table 4
Agreement rate per emotion domain between two and three annotators.

Number of annotators agreed	Trust	Disgust	Anger	Fear	Anticipation	Joy	Surprise	Sadness
2 annotators	100%	50%	80%	57%	100%	100%	100%	100%
3 annotators	50%	0%	20%	29%	100%	67%	100%	100%
Number of labels	205	24	9	79	37	27	6	11

that we needed 43 receivers in each of the six combinations of sender–receiver pairs (i.e., Poland, US, LLM). Therefore, Polish and US sender groups had to write 43 emails each. Since each sender group sends only a small number of emails, it is difficult to determine whether the differing responses from two receiver groups are due to specific characteristics of an individual sender's email read by a particular group, or if they reflect broader cultural or linguistic differences among the receiver groups. To mitigate this issue, we have implemented a control measure where receivers from both countries will receive the same set of emails. This means that Polish and US groups of senders will write a batch of 23 emails each (i.e., 46 unique emails will be written in total), and each batch will be replicated and sent to both US and Polish receivers.

For LLM-generated emails, we produced 1 unique email[5] per use case for each type of LLM (i.e., $LLM_{generic}$ and $LLM_{emotion-aware}$). For distributing both human-generated and LLM-generated emails to receivers, we first replicated the unique emails to ensure that the statistical power was maintained, and then we randomly assigned them to the two groups of receivers (i.e., Poland and US). In total 266 Polish and US receivers obtained 84 emails from Polish senders, 93 from US senders, and 89 written by the LLM. Out of LLM-generated emails, 43 emails were written by $LLM_{generic}$ and 46 by $LLM_{emotion-aware}$.

4.4. Emotion label mapping

To synthesize and standardize emotion annotations done by receivers, we mapped all emotion labels to Plutchik's Wheel of Emotions Plutchik (1980), a widely used taxonomy in emotion research that captures eight primary emotion categories: joy, trust, fear, surprise, sadness, disgust, anger, and anticipation.

The annotation process involved three independent researchers who categorized each emotion label provided by participants into one or more Plutchik categories. When necessary, labels were mapped to dyads (combinations of two adjacent Plutchik emotions) if the affective nuance was better captured in that way.

Agreement was distributed as follows: in 36% of cases all three annotators agreed, and the label was directly retained; in approximately 45% of cases two annotators agreed while one disagreed, in which case the majority label was adopted (bringing the total pairwise agreement to about 81%); and in the remaining 19% of cases all three annotators disagreed, where consensus was reached through discussion. Previous research has shown that agreement tends to be lower for nuanced or interpersonal communication contexts due to inherent ambiguity in emotional expression Devillers et al. (2005). Even when we binarized Plutchik's emotions (i.e., computed the distribution for positive emotions *vs.* the negative ones, we obtained an agreement of 60% for positive and 19% for negative between three annotators. The agreement rates and number of labels per emotional domain are presented in Table 4, and the total number of labels is greater than the number of emails as we allowed a receiver to assign several emotions to the email, such that one email could have multiple labels. The highest agreement rates were on the positive emotional domains (trust, anticipation, joy), and the lowest on the negative domains (disgust, anger).

[5] By limiting the output to one email per case, we intended not to introduce any other factors to our study design related to the model's parameters. However, we initially experimented with generating multiple versions of the same email, which resulted in only slight textual variations.

4.5. Dataset

After collecting the data, we performed three post-processing steps: *a)* conducted quality checks and filtering, primarily aimed at removing AI-generated answers and correcting spelling errors; *b)*collected responses with the demographic information provided by Prolific; and *c)* mapped the emotions labeled by humans to the eight emotional categories according to the Plutchik's wheel of emotions (Plutchik, 1984): trust, fear, surprise, sadness, disgust, anger, anticipation, and joy. The emotion annotation was conducted by three independent researchers, who initially agreed on 36% of the cases. They resolved disagreements through collective discussion to reach a consensus. The procedure and mapping are described in the Appendix.

The final dataset contained 266 observations for the variables "task clarity", "willingness", and the type of email "isLLMGenerated" (i.e., a binary variable indicating whether the email was human-generated or LLM-generated; 1 indicates LLM-generated email, and 0 indicates human-generated email). Additionally, it contained the following variables: "origin combination" indicating the sender's and receiver's origin (i.e., US-US, US-Poland, Poland-Poland, Poland-US, $LLM_{generic}$-US, $LLM_{generic}$-Poland, $LLM_{emotion-aware}$-US, $LLM_{emotion-aware}$-Poland); the "type" indicating whether the "sender" was "human" $LLM_{generic}$ or $LLM_{emotion-aware}$; the "power equality" between the sender and receiver (1 indicates same power level and 0 indicates different level); and a set of binary emotional variables: "trust", "fear", "disgust","surprise", "sadness", "anticipation", "joy", "anger".

5. Results

Before conducting any analysis, we examined the normality of our variables (i.e., task clarity and willingness) using Shapiro-Wilk normality test, and found that both were not normally distributed (task clarity: $W = 0.92$, $p < 0.001$; willingness: $W = 0.95$, $p < 0.001$). Therefore, we used a non-parametric Mann-Whitney U test for comparing two groups of continuous variable values and a non-parametric Random Forest model for fitting regression of continuous variables versus the set of categorical variables. Additionally, due to multiple tests being applied to the same data, we adjusted the p-values using Bonferroni correction.

RQ_1: Do human-generated emails differ from LLM-generated ones in terms of task clarity and willingness to act upon an email?

LLM improves clarity of the emails. We compared the average task clarity values for both human-generated and LLM-generated emails. The LLM-generated emails performed better, with an average score of 72.9% (SD = 27.8), for a combination of "$LLM_{emotion-aware}$" and "$LLM_{generic}$". In contrast, human-generated emails scored only 60% (SD = 20.3). We found a statistically significant difference in task clarity between LLM-generated and human-generated emails ($p < 0.007$ after Bonferroni correction), with the absolute rank-biserial correlation effect size of 0.13. The distributions of scores are shown in Fig. 3.

LLM does not improve the willingness to act upon the email. The average willingness to act upon LLM-generated emails was 69%, representing a marginal increase of 10% over the average scores for human-generated emails which were 63%. Standard deviations of both groups were close to each other, with a value of 25% for the human-generated emails and 24% for the LLM-generated ones. There was no statistically significant difference in willingness to act between LLM-generated and human-generated emails ($p = 0.05$ after Bonferroni correction).

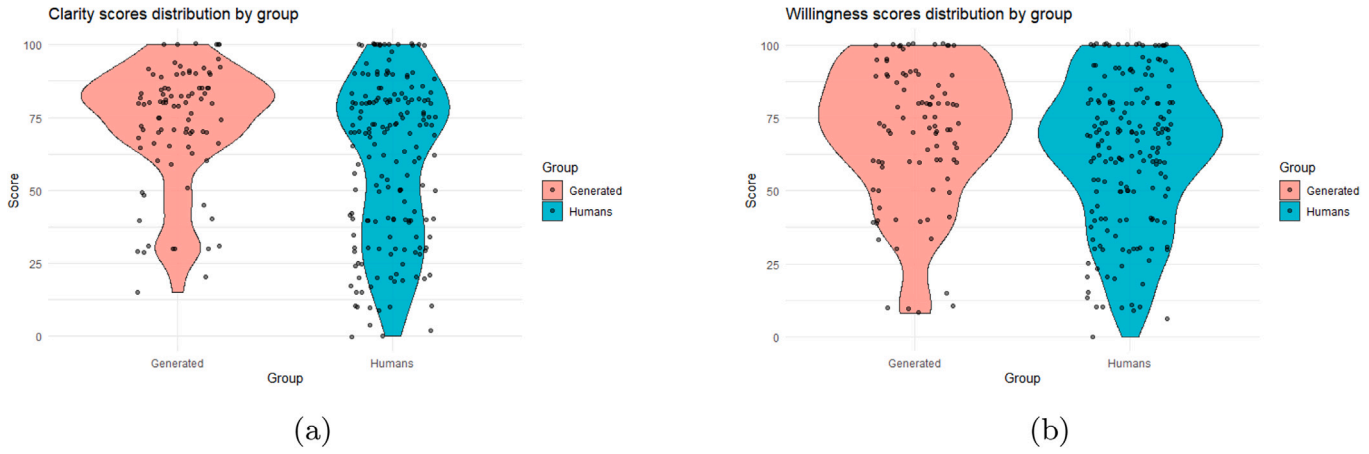


Fig. 3. Distribution of the clarity (a) and willingness (b) scores for generated and human-authored emails.

Summary: As one expects LLMs have indeed improved the clarity of emails. In fact, a previous study highlighted that generative AI improves the clarity in academic writing (Santiago Jr et al., 2023). This can be explained by the capabilities of LLMs to produce fewer grammatical errors, while maintaining the same vocabulary as human writers (Lee et al., 2022), or to produce texts that are more complete, cohesive, and coherent than those written by humans (Tirado-Olivares et al., 2023). However, we did not find any evidence that LLMs moved our participants to action. To test which factors may influence people's willingness to act upon LLM-generated emails, we tested three hypotheses.

RQ_2: What is the role of cultural background, power differences, and perceived emotions in people's willingness to act upon emails, and to what extent do LLMs' prompting strategies relate to the impact of these factors?

H_1: *The willingness to act upon an email increases if sender and receiver share the same cultural background.*

Previous studies found that people from different cultures share information with each other in different ways (He et al., 2010). For example, East Asians generally tend to consider the group's needs and viewpoints in their communication (sociocentric behavior), while Pakeha are more focused on individual needs and self-expression (idiocentric behavior) (Pekerti and Thomas, 2003). Building on these findings, one may hypothesize that, in our case of email communication, when both the sender and receiver of a message share the same cultural communication style—be it sociocentric or idiocentric—they might more effectively persuade one another. As a result, this alignment could potentially lead to a higher willingness on the part of the receiver to perform a task, as compared to scenarios where their communication styles differ. This hypothesis draws on the idea that shared communication styles based on culture may enhance understanding and motivation to act upon a task.

To test this hypothesis, we conducted a Mann-Whitney U test on a dataset slice that contains only human-generated emails. We did not include LLM-originated emails in this test as even though largest LLMs are trained predominantly on English corpora, training corpus usually contains data in other languages (an example would be disclosures from LLaMA development team Touvron et al., 2023), so we cannot label LLM emails as emails originated by a single culture. Therefore, we compared two types of emails: one where the sender and receiver are from the same country (same-country group), and another where the sender and receiver are not from the same country (non-same-country group). The average willingness to act for the same-country group was 67.3%, which was higher than for the non-same-country group (58.5%). However, after adjusting the p-value with Bonferroni correction, Mann-Whitney U test showed no statistical difference in the willingness to act, regardless of the sender and receiver's country of origin combination ($W = 3122.5, p = 0.02$). Therefore, we reject H_1.

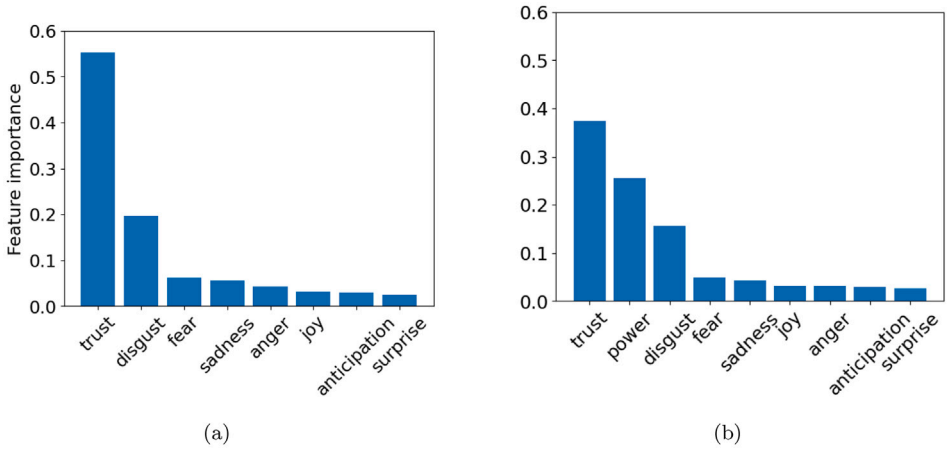
H_2: *The willingness to act upon an email increases if the recipient is subordinate.*

Previous research has shown that an individual's awareness of their own high power level can decrease their motivation to engage in routine and uninteresting tasks (DeWall et al., 2011). Additionally, studies indicate that groups with significant power disparities tend to perform better when the member with the highest power is also the most competent (Tarakci et al., 2016). Based on these insights, we hypothesize that if the sender holds a higher power level than the receiver, the receiver may be more inclined to complete the task, potentially leading to enhanced task completion efficiency within the group.

To test this hypothesis, we analyzed the full dataset of emails (both human- and LLM-generated). We defined power levels strictly based on use cases' descriptions, ensuring that neither human nor LLM "senders" creating the emails could modify these definitions. Among the four use cases, we had one case with the same power level between the sender and receiver, and three cases with different power levels (i.e., a manager wrote an email to a subordinate). This allowed us to test if the receiver's willingness to act upon the task differed in these two situations. Upon comparing the two groups (i.e., one with the same power level and the other with different power levels), we found a statistically significant difference in the willingness to act ($p < 0.007$ after Bonferroni correction). This means that if the sender is of a higher power level, the receiver is more willing to act (70%) on average than in the case when both sender and receiver have the same power level (54.6%). Therefore, we accept H_2.

H_3: *The willingness to act upon an email increases if the receiver perceives positive emotions in the email.*

Previous works explored how various factors (e.g., competency, benevolence, communication quality, integrity, and consistency) affect the trust level at work (Sekhon et al., 2014). Kiuru et al. (2020) demonstrated that the emotions a worker feels at the start of a task can influence both the amount of effort they put into the task and the eventual outcomes. Brundin et al. (2008) found that managers' appreciation and expression of satisfaction in a project propagate positively to their subordinates. In contrast, expressing frustration has the opposite effects. Therefore, one may hypothesize that the emotional context set by a manager at the outset of a task (in an email) significantly impacts employees' motivation and behavior. Positive emotional expressions (e.g., satisfaction) may increase employees' willingness to take initiative and act upon tasks. Conversely, negative expressions (e.g., frustration) may decrease their willingness to engage actively and act upon tasks.



(a) (b)

Fig. 4. Random forest models' feature importance on: *(a)* eight emotions (trust and disgust were the most important features); and *(b)* eight emotions and power level (trust and power level were the most important features).

To test this hypothesis, we used the full dataset of emails (both human- and LLM-generated) and conducted two tests: one to check the difference between willingness to act upon a task for groups of receivers perceiving positive and negative sentiment while reading emails, and another to identify which emotion categories affect receivers' willingness to act upon the tasks expressed in the emails.

By running a Mann-Whitney U test for two groups of receivers experiencing positive and negative sentiment, we found a statistically significant difference between the two groups ($p < 0.007$ after Bonferroni correction). On average, we found that the receiver's willingness to act upon the task increased by 24% in case of perceived positive emotions (71.7%) compared to negative ones (58%). To identify which emotion categories affect receivers' willingness to act upon emails, we fitted a Random Forest model. Unlike previous non-parametric tests, this model can help in identifying the relative importance of each emotion, allowing us to determine the most influential one(s). Our model explained 17.14% of the variance (OOB = 520.1), with the most predictive emotions being "trust" and "disgust" (Fig. 4(a). This is further supported by the fact that the median willingness to act for emails labeled as "trust" was 72% and for "disgust" was 39.5%. Therefore, we accept H_3.

Given that these two factors are important independently, we then tested them in conjunction. By fitting another Random Forest model with both the emotion categories and the power level, the model explained 24.68% of the variance (OOB = 472.7) with the most predictive features being the trust and the power level (Fig. 4(b). This means that whenever an email conveyed trust and the sender held a higher power level than the receiver, the receiver was more willing to act upon it.

Apart from culture, power and emotions, we have explored impact of education level and age of senders and receivers on willingness to act and clarity. For that, we fit a random forest regression model for willingness versus education level and age of senders and receivers with only human-generated emails (as we cannot confirm LLM notion of age and education). The model was trained using 500 trees, with 1 variable tried at each split. The model explained only 0.03% of the variance (OOB = 708.7), indicating poor predictive performance. Cross-validation results further confirmed the lack of predictive power, with R-squared values close to zero and high error metrics (e.g., RMSE = 25.22, MAE = 20.57 for 'mtry = 2'). These results suggest that the sender's age and education level do not provide meaningful information for explaining receivers' willingness to act and perceived clarity of the email.

Qualitative analysis of emails from the most impactful emotional categories. Upon reviewing the emails across the range of willingness to act scores, we found that 113 out of 126 emails in the two upper quartiles were labeled with trust, while only two had a disgust label (with no disgust labeled emails in the 4-th quartile). Conversely, in the two lower quartiles, 20 out of 127 emails were labeled with disgust (with 14 emails in the 1-st quartile). Overall, the dataset contained 24 disgust labels and 204 trust labels. This indicates that nearly all the least successful emails were associated with disgust, while the majority of the most successful emails were linked to trust. Given that these two emotions were previously identified as the most influential predictors of willingness to act, we thematically analyzed the emails in the two bottom quartiles associated with disgust and those in the two top quartiles associated with trust.

Thematic analysis was conducted using a combination of open and axial coding (Clarke and Braun, 2021). Next, we discuss the most prominent themes, and support each theme with phrases identified in emails. When interpreting these results, remember that we mapped[6] the emotion labels to Plutchik's basic emotional categories (Appendix).

For emails labeled with 'trust', we identified the following themes: "acknowledgment of receiver's expertise", "expression of gratitude and appreciation to receiver", "acknowledgment of receiver's time value", "offering support and clarification for the task", "trust in receiver's abilities". The first theme ("acknowledgment of receiver's expertise") is about the use of positive words towards the receiver as well as receiver's past successful deliverables (e.g., *"I've heard great things about your previous work"; "Your opinion is very important to me, because as a security policy specialist, you will help me…".*) The second theme ("expression of gratitude and appreciation to receiver") highlights the importance of acknowledging the receiver's effort and potential contribution to the task (e.g., *"Both the client and I appreciate any assistance you can provide."; "I appreciate your help in this matter and look forward to your contributions to the spreadsheet."*). The third theme ("acknowledgment of receiver's time value") emphasizes the sender's recognition of the receiver's busy schedule and the importance of their time (e.g., *"Please let me know your availability and I will be happy to work around your schedule"; "I understand that you have a lot on your plate"*). The fourth theme ("offering support and clarification for the task") focuses on the sender's readiness to assist and provide detailed explanations, particularly when the task involves complex technical elements (e.g., *"Please reach out with any questions."; "I hope everything is clear; if not, please do not hesitate to ask."*). Finally, the fifth theme ("trust in receiver's abilities") is about expressing confidence in the receiver's skills and competence to handle

[6] "Trust" included the labels: respectful, supporting, encouraging, polite, understanding, assertive, directive, formal, flattering, and highly expectant. "Disgust" included: neglecting, blaming, pretentious, in a hurry, annoyed, hurried, brief, and rushed.

the responsibilities assigned to them (e.g., *"I am confident that you will handle these responsibilities efficiently"; "I trust in your ability to convey this information accurately and promptly"*).

For emails labeled with 'disgust', we identified the following themes: "neglecting receiver's time", "deciding upon the receiver", "blaming", "machine talk", "using idioms or jargon". The first theme ("neglecting receiver's time") relates to situations where the sender indirectly suggests that the receiver's time is less valuable (e.g., *"I don't have much time to study the product—and to be honest I need your help.")*. The second theme ("deciding upon the receiver") involves senders making assumptions or decisions about the receiver's time and capacity to complete tasks. It is characterized by senders dictating how long tasks should take or whether the receiver is able to undertake them, regardless of their actual capacity or existing workload (e.g., *"I don't think this will take a lot of your time"; "I understand that you have a lot on your plate, but this is a high priority task".*) The third theme ("blaming") is about situations where the sender blames their own team, the receiver or a situation (e.g., *"The project deadline was committed without considering the learning curve for a new product like this"; "The team expended more time on our current project than anticipated"*). The fourth theme ("machine talk") is identified in emails that deliver a factual, straightforward description of the task without offering any rationale for its necessity or acknowledging the receiver's time or expertise. This type of communication often results in receivers feeling undervalued (e.g. *"Can you download the cost info?… By the way, you'll need to adjust the extract parameters"*). Finally, the fifth theme ("using idioms or jargon") involves emails characterized by the use of specialized language or complex expressions that can come across as pretentious (e.g., *"I have done due diligence"; "jeopardy"; and "efficiently iron out any potential issues")*.

6. Discussion

6.1. Main findings

We collected 46 unique human-generated and 8 LLM-generated emails, and presented them to 266 participants who assessed the clarity of conveying a task and the extent to which participants were willing to act upon it. Consistent with prior work, we found that LLM-generated emails were rated as clearer than human-written ones. However, this increase in clarity did not lead to higher willingness to act on the emails' requests. Instead, our data show that willingness to act was primarily influenced by two factors: *i)* the emotional tone of the message as perceived by the reader; and *ii)* the power relationship between sender and receiver. Contrary to expectations, cultural alignment between sender and receiver had no significant effect.

While our findings on clarity confirm prior evidence that LLMs can improve the readability of workplace communication, the novelty of this study lies in going beyond surface-level efficiency. By disentangling how cultural background, power asymmetries, and perceived emotions influence recipients' willingness to act, we highlight that social dynamics (rather than LLM use alone) are the decisive determinants of cooperation in email communication.

6.2. Implications

Our work contributes to advancing our understanding of culture, power dynamics, and emotions in email communications, and seeks ways to improve them using LLMs.

In terms of cross-cultural communication, we found that even if the sender and receiver had different communication styles (Meyer, 2014; Hofstede, 2011), it did not significantly change the willingness to act upon the task. This observation aligns with previous work, which found that individuals with different communication styles could find common ground and contribute equally to decision-making processes, often by adapting to each other's cultural norms (Zakaria, 2017).

The strongest predictor of willingness to act was perceived emotional tone. Emails labeled by readers as expressing supportive emotions (e.g., encouragement, appreciation, and politeness) were associated with higher willingness to comply. By contrast, emails perceived as disrespectful, demanding, or blaming were more likely to generate reluctance or refusal, regardless of whether the sender had higher formal power. Emails from senders perceived as higher in the organizational hierarchy (e.g., "manager" roles) generally increased willingness to act; however, this effect was amplified or suppressed by the emotional tone. In several cases, messages from high-power senders triggered emotions such as fear or disgust, especially when they conveyed urgency or blame. This shows that power is not inherently persuasive: its effect depends on whether the message maintains interpersonal respect. Ignoring receiver's potential emotional reaction accounts for up to 17.4% of the variance in the receiver's willingness to act. This finding is consistent with prior work, which reported a correlation between management's perception of future success and employees' willingness to engage in entrepreneurial behaviors (Brundin et al., 2008). Our design minimized confounds by providing receivers with no prior information about relationship history. Thus, any "trust" perceived was inferred from the email's tone and phrasing, and not from assumptions about long-term relational context. This supports the interpretation that emotional perception drives behavioral response in mediated communication.

Although LLM-generated emails scored significantly higher in clarity, this did not translate into higher willingness to act. This finding echoes earlier concerns in the literature that syntactic clarity does not guarantee interpersonal alignment Hohenstein et al. (2023a). Emails that were technically clear but emotionally flat, impersonal, or overly directive often failed to persuade recipients. In contrast, human-written emails that embedded emotionally nuanced language (e.g., gratitude, humility, acknowledgment) were more persuasive despite being less structured. This highlights a key insight: effective communication combines clarity with emotional intelligence. LLMs, though excellent at clarity, currently struggle to model affective nuance unless explicitly prompted to do so.

Our thematic analysis of the emails also provides actionable insights into how specific linguistic choices shape emotional perception. When crafting the use cases for our study, we intentionally injected potential emotional triggers (e.g., "However, you have requested Mike for estimations one time before for another project, and your team spent more than he estimated, so you had to implement contingency measures to meet the project cost targets and your team missed bonus due to poor financial results"—Case 3). This allowed senders to decide whether they would react to these triggers and include relevant information in their emails, or ignore them altogether. For example, senders who included information from the positive emotional triggers in their emails generally succeeded in persuading receivers to act on the tasks. Overall, we identified that the major factors enabling receivers to perceive support, encouragement, and respect include explicit acknowledgment of the receiver's expertise and importance, expressions of gratitude, acknowledgment of the receiver's valuable time, and offers of support or clarification regarding the task. On the other hand, emails that incorporated information from the negative emotional triggers, such as references to past failures or external blame, tended to fail in persuading the receiver. We also found that expressions of appreciation can inadvertently become neglectful if the sender acknowledges the receiver's qualifications while disregarding the importance of their time. Furthermore, we observed that emails lacking clear expressions of appreciation, gratitude, and respect for the receiver's time were perceived as neglectful, despite not containing any obvious expressions of disgust. These findings illustrate that crafting a persuasive email is not straightforward. It requires a multifaceted approach that involves removing any toxic or unnecessary elements and then incorporating the necessary positive elements to enhance the receiver's willingness to act.

Large Language Models (LLMs) have demonstrated significant potential in enhancing written communication, particularly in workplace settings. Research shows that these models can substantially improve writing efficiency and creativity, especially benefiting those with lower skill levels (Dell'Acqua et al., 2023) or novice users, such as law students

working on persuasive legal cases (Weber et al., 2024). While our study confirms LLMs' ability to enhance email clarity, the relationship between AI and communication effectiveness presents a complex dynamic.

Previous research has highlighted a dual impact: while AI tends to improve message structure and positivity, recipients may react negatively to communications they perceive as AI-generated (Hohenstein et al., 2023a). Studies have shown that although LLMs can achieve persuasive outcomes comparable to human efforts, their methodologies differ, potentially both enhancing and compromising informational integrity (Carrasco-Farre, 2024). Nevertheless, our findings suggest that the benefits of LLM-generated emails generally outweigh the potential drawbacks of AI perception, particularly in terms of clarity improvement.

However, an important limitation emerged: LLM-generated communications did not necessarily improve recipients' willingness to act. This aligns with research by Meng and Dai (2021), which found that AI chatbots expressing emotions without offering genuine support could actually increase client stress levels, unlike human consultants who successfully reduced them. This is consistent with our observation that "offering support" was prevalent in emails with the highest willingness to act.

These findings have practical implications for AI interface and prompt design. While most LLM tools default to clarity-focused output, this approach may be insufficient for emotionally sensitive or persuasive communication. Recent work on "blended work" (Constantinides et al., 2025) emphasizes that writers co-creating with AI must actively guide tone and interpersonal framing, with even minor prompt adjustments potentially causing significant changes in emotional tone. Our research suggests a concrete path forward: incorporating trust-related emotional domains in training prompts may enhance recipients' willingness to cooperate.

As we continue developing AI assistants, it is crucial to ensure they incorporate these key positive themes to enhance message effectiveness and improve receiver engagement. This requires moving beyond mere clarity to include explicit expressions of support, respect, and empathy through carefully designed prompt scaffolds or affective templates.

6.3. Limitations and future work

Our study has seven main limitations that call for future research efforts. First, the study was conducted with participants from only one pair of countries (i.e., US and Poland), which may limit its generalizability to other cultural contexts. Future work should replicate our approach with additional country pairs to better capture cultural variation. Second, while we aimed to account for cultural and power dynamics through scenario design, we did not include explicit manipulation checks. Receivers were not told about the sender's cultural background, and role-based cues (e.g., "manager" vs. "team member") were used to signal power asymmetries without directly validating how receivers perceived them. Future studies should incorporate direct manipulation checks to ensure these factors are interpreted as intended. Third, our treatment of emotion focused solely on receiver-perceived affect, which we consider more relevant for explaining willingness to act; however, this means that sender-intended emotions were not captured. Moreover, mapping open-ended emotion labels to Plutchik's taxonomy involved subjective judgment, with modest initial inter-annotator agreement (36%). Although disagreements were resolved through discussion, future studies may benefit from dimensional approaches (e.g., valence–arousal) to improve reproducibility. Fourth, our sample was restricted to knowledge workers with at least a college degree in technology, telecommunications, or finance, and we assumed they would comprehend the generic, non-industry-specific scenarios without time pressure. A more refined approach would be to conduct similar studies within single organizations, recruiting participants with homogeneous expertise. Fifth, we relied on one specific LLM (GPT-4) at the time of the study. Since then, new models (e.g., Claude, Gemini, Llama 3, and Mistral) have emerged, and replication across multiple systems would strengthen the robustness of our findings. Sixth, our measure of willingness to act relied on a single self-reported item, which may not fully capture the ecological validity of workplace compliance. Future research should incorporate behavioral or longitudinal measures (e.g., tracking actual task completion) to strengthen external validity. Finally, we tested writing emails purely by humans and purely by AI, but did not explicitly test email co-creation. However, checks with the GPT Zero tool.[7]

7. Conclusion

We conducted a crowd-sourcing study with 266 participants to understand the role of LLMs in email communication in terms of task clarity and willingness to perform the task expressed in an email. We found that LLM-generated emails expressed tasks in a clearer way but performed comparably to humans for calls to action. Cultural and emotional concordance between the sender and receiver was not important; however, the disparity in power levels and the emotions perceived by the sender were the most influential factors that moved receivers to take an action.

CRediT authorship contribution statement

Marina Polupanova: Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation. **Marios Constantinides:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Conceptualization. **Daniele Quercia:** Writing – review & editing, Supervision, Project administration, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A

A set of 4 use cases was constructed by following best practices for project management according to the Project Management Body of Knowledge book (PMBOK 7th edition). The four use cases were about data gathering and analysis (Case 1 and Case 2), estimating (Case 3), and meetings and events (Case 4). Each case included a task description such as providing a spreadsheet with costs or a Word document with system details. Additionally, the cases outlined the sender's and receiver's positions and qualifications—for example, the sender might be an experienced Solution Architect, and the receiver could be a project manager. Moreover, each case incorporated a potential emotional trigger for the sender such as the receiver's past performance on a specific task, to gauge the emotional dynamics involved in the communication.

A.1. Use cases descriptions

Case 1: You are a solution architect for a large software integration project. You have your own checklist for data gathering for the new project, but it turns out that one of the customer requirements will demand integrating a product FeaturedProduct which you have never used before in your design, so respective data placeholders are not present in your checklist. As the project deadline was committed without accounting for the learning curve for anything new, you realize you have almost no time for a complete product training. You have gotten another contact of the solution architect, Paul, who may clarify your doubts as he has already worked with that product in the past for GemY company project, or even sit with you on the call to discuss your questions. You have heard from the GemY team that the deployment using Paul's design went smoothly. You downloaded the product description and basic design guidelines from the company product documentation repository,

[7] https://gptzero.me/.

but you will need to figure out if there are any risks or limitations that you need to consider or extra integration steps to be done for other design components with certain customer requirements descriptions.
Task. Please write 150–250 word request e-mail to Paul using everything you know from this case description.
Case 2: You are the manager of the project, and you want to run an extra evaluation of your project performance in terms of cost as you suspect that your team has spent more than planned for completed activities due to a known hourly rate increase this month. To be able to create the corresponding report, you will need to ask your team member, Lisa, to download the actual cost information for you from ERP (Enterprise Resource Planning), and then compare it to your initial cost plan. The default table, which can be downloaded from ERP does not contain all the information needed for report, so you will need to adjust ERP extract parameters to ensure that the downloaded file you receive contains each team member's ID, their hourly cost for certain month and the number of hours booked. For comparison of planned and actual spending, you want to receive a separate summary table. As of now, the team has completed first five tasks from the project plan and another five are left to be done. You will need to see in summary a comparison of planned and spent costs and days for the first five tasks, while last five tasks will contain only planned data.
Task. Please write 150–250 word request e-mail to Lisa using everything you know from this case description.
Case 3: Your company is bidding for a consulting data analytics project for a customer "Alltime". After reading customer requirements you want to request the help of Heather and Mike, experts from your company, in estimating the workload and list of tools needed for the new project. You believe that analogous estimating will be the best method fitting the current case, so that activities done in similar previous projects would be used to approximate workload of current project activities. However, you have requested Mike for estimations one time before for another project, and your team spent more than he estimated, so you had to implement contingency measures to meet the project cost targets and your team missed bonus due to poor financial results. To increase planning precision this time, you are creating the spreadsheet in the shared folder, where you add not only the columns for the workload value and the required tools for each work item, but also "past experience" column where you want to populate the name of the project where such service was performed and its major risks and resulting workload. These "past experience" entries should be descriptive enough to justify why it is sufficient to use that past record for estimating tasks in new project.
Task. Please write 150–250 word email to Heather and Mike, asking them to populate the workload estimation spreadsheet, using all that you know from this case description.
Case 4: You are a Program Manager and oversee the execution of multiple projects for customer, ClearWater. A Project Manager, Bill, assigned to a new Security Policy update project within your Program several days ago, sent you an email asking about the frequency and content of the reports he needs to send you regarding the project's progress. You have not worked with Bill before, but you have heard excellent feedback about him from his previous manager. You have regular Program meetings with this customer every Wednesday, so you must include slides on this project in your overall slide deck at least one day before the customer call. Your deck on all Program projects includes their short scope and deliverables description, a chart with the project progress, and a list of current risks and issues. You also describe the proposed mitigation of all indicated risks during the customer calls. As the customer can ask very detailed questions during the discussion and expects most of the answers right at the meeting, it always helps to have a clarification call on project details with every Project Manager in your group regarding his summary prior to customer call.
Task. Please, write 150–250 word response email to Bill, answering his question based on all that you know from this case description.

A.2. Mapping of the emotion labels to plutchik's basic emotional categories

During the processing of the data, we mapped answers of the readers to Plutchik's basic emotional domains (in case the emotion better fits the dyad, it will be recorded as part of two emotional domains, creating the dyad). The annotation was carried out by three independent researchers who achieved an initial agreement in 36% of the cases. In 14% of the cases, where each researcher initially provided a different label, the discrepancies were resolved through discussion to reach a consensus.

- **Trust**: "Respectful", "Supporting", "Encouraging", "Polite", "Understanding", "Assertive", "Directive", "Formal", "Flattering", "Highly expectant"
- **Anticipation**: "Curious", "Aggressive/pushy"
- **Disgust**: "Neglecting", "Blaming", "Pretentious", "In a hurry", "Annoyed", "Hurried", "Brief", "Rushed"
- **Joy**: "Excited", "Normal","Grateful"
- **Anger**: "Frustrated", "Blaming", "Aggressive/pushy", "Insistent", "Demanding", "Pushy"
- **Surprise**: "Surprised"
- **Sadness**: "Unhappy"
- **Fear**: "Worried", "Hesitant", "in Panic", "Embarrassed", "Nervous", "Confused", "Cautious"

A.3. Examples of the generated emails

Example emails generated by participants from Poland.
Email: Hi Bill! Thanks for your email regarding my expectations as to scope and frequency of your reports regarding the ClearWater projects. I believe it makes a lot of sense to make this expectations clear from the get-go rather than rely on guesswork. As to how often I'm going to need updates from, I run regular weekly meetings with this customer and am expected to provide them with overviews on all relevant areas. I put presentations together based on input from distinct areas and am going to need yours every week on Mondays by noon; if for any reason significant stuff is bound to happen in the first half of the week that you're not able to report by then, please let me know in advance, too. As to the scope, I need to include all relevant SecurityPolicy deliverables, as well as material to include in the project progress chart and a list of current risks and issues you're encountering and your ideas for risk mitigation. The customer tends to ask quite detailed questions in our meetings, so I'll be contacting you if I decide I need to get more deeper on specific matters. Please let me know if there's anything else I can help you with. Best regards.
Email: Hi Paul, I'm contacting you regarding your previous work for the GemY team on the FeaturedProduct. I have been hired as a solution architect for a project integrating FeaturedProduct. As far as I'm concerned deployment using your design went smoothly, so I would love to get to know your perspective. I have already downloaded the product description and basic design guidelines from the company product documentation repository, but I will need to figure out if there are any risks or limitations which I need to consider or extra integration steps to be done for other design components having certain customer requirements description. I would greatly appreciate it if you could spare me sometime tomorrow, I suppose 1h or 1.5h would be enough for me to pick your mind on this topic. Sincerely, XYZ.
Email: Hi Heather and Mike! I hope you're well experts. I would like to ask you something - Could you populate the workload estimation spreadsheet? I need your help again. Mike, last time when you helped me with another project, I had to implement contingency measures to meet the project cost targets and my team missed bonus due to poor financial result. I would like to increase planning precision. In order to do that I decided to create the spreadsheet in the shared folder. I will add columns for the workload value and tools required for each work item.

Please help. This past experience entries should be descriptive enough to justify why it is enough to use that past record for assessing task in this project. I'm counting on you. I hope this time everything will be ok, and we will achieve good financial results because we are more experienced, thank you in advance!

Example emails generated by participants from US.

Email: Hi Paul, We have an upcoming project using the product FeaturedProduct and I was hoping to utilize some of your expertise from when you worked with it at GemY. This is a new product in my scope, and while I have some training materials and guidelines, I would love to schedule a call with you sometime this week to go over my notes and any questions I have regarding the product. The aspects I am most concerned with include the risks/limitations needing to be considered and any additional integration steps needing done based on my client's requirements. I have heard from your previous team at GemY that you are a subject matter expert with FeaturedProduct, so I would really appreciate any assistance you can provide.

Both the client and I appreciate any assistance you can provide. Please let me know what works best with your schedule over the next few days and I can set up a call to review.

Thank you!

Email: Hi Heather and Mike, I was hoping that you both could add your valuable experience to a project that I have begun to work on. Your experience on previous projects has proved very insightful. I'm working on submitting a bid for the Alltime project and would like to lean on the overhead data from previous projects in order to create a better estimate of workload value and risks. I've created a shared spreadsheet that we all have access to in order to more easily facilitate the transfer of this information. More specifically, I'm hoping to gain your assistance in populating the "past experience" column. Previously, I've relied too heavily on the exact data of experience and would to like to take a broader approach when outlining the bid for Alltime. The data I need is the work done on the previous project, and how much workload was generated by servicing the project. Another aspect to outline are the major risks to the project, and how reliable using this estimate would be for the "Alltime" project. Thanks for your time.

Email: Hi Bill, I hope this email finds you well. As you probably already know, I have my update meeting with the customer on Wednesday of every week. Would it be possible to provide your portion of the program such as your status reports, on Friday of every week? This would allow me to review them on Monday morning, and then we could have a quick call Monday afternoon for any questions I may have or clarification I may need on that weeks details. I can then compile them into my weekly presentation on Tuesday and be well versed with your information in preparation for the Wednesday presentation. The customer often asks very detailed questions, so I would much rather be overprepared than underprepared. Let me know if this works for you. As always, I appreciate your hard work and dedication. Regards.

Example emails generated by LLM (generic).

Email: Dear Paul, I hope this message finds you well project that requires the integration of FeaturedProduct. I understand that you have previously worked with this product on the GemY company project and that the deployment went smoothly. Given your experience, I was hoping you could provide some guidance. Due to a tight project deadline, I am facing a challenging learning curve with this new product. While I have downloaded and reviewed the product description and basic design guidelines from the company product documentation repository, I am concerned about potential risks, limitations, and extra integration steps that may not be explicitly highlighted in these documents. Since you have practical experience with this product, could you help me understand these aspects? Specifically, I am keen to know if there are any additional considerations when integrating FeaturedProduct with other design components, given certain customer requirements. I realize you are probably busy, but I believe your insight would be invaluable in helping me achieve a successful deployment. If you are available, we could arrange a call at a time that suits your schedule. I appreciate your time and look forward to hearing from you soon. Best regards, Andrew.

Email: Dear Lisa, I trust this email finds you well. I am writing to request your assistance in conducting an additional assessment of our project's cost performance. I have some concerns that we might have exceeded our planned costs for the tasks completed due to the recent rise in hourly rates.

Could you please extract the actual cost information from our ERP system? Please ensure to adjust the ERP extract parameters so that the downloaded file includes each team member's ID, their hourly cost for this month, and their recorded hours.

Alongside this, it would be beneficial if you could prepare a summary table to compare our actual costs and the initial cost plan. As you know, we have completed five tasks from our project plan, and another five are still pending. The summary should reflect the planned and actual costs and days for the first five tasks, while the last five should only contain the planned data.

This evaluation will help us understand our spending patterns better and ensure that we stay on track with our budget. I understand that you have a lot on your plate, but this is a high priority task, and I would appreciate it if you could provide the report by the end of this week.

Thank you for your support in this matter. Best Regards.

Email: Dear Heather and Mike, I hope this email finds you well. As you might be aware, we are currently bidding for a data analytics project for our prospective client, Alltime. Their requirements necessitate an in-depth understanding of both the workload involved and the tools required for the new project.

I have created a spreadsheet in the shared folder for us to collaborate on these estimations. It includes columns for the workload value, required tools for each work item, and an additional column titled "Past Experience." In this column, I would like you to detail the name of the project where a similar service was performed, its major risks, and the resulting workload.

The aim is to employ an analogous estimating approach to predict the workload of current project activities using data from similar past projects. Your input is crucial to ensure we adequately capture the scope of the project and plan appropriately.

I understand that our previous project estimation was underspecified, which led to a need for contingency measures. This time, I would like us to be as thorough and precise as possible to avoid any such discrepancies.

Please make sure your entries in the "Past Experience" column are descriptive enough to justify why it is adequate to use that past record for estimating the task in the new project.

I appreciate your help in this matter and look forward to your contributions to the spreadsheet. Best regards, Andrew.

Example emails generated by LLM (emotional).

Email: Dear Paul, I hope this message finds you well. I am writing to you as I am currently working on a software integration project where I am required to integrate a product called FeaturedProduct. I am quite familiar with the majority of the project specifics, but this product is new to me, and I have not had the chance to use it in any of my past designs. I have heard great things about your work with FeaturedProduct on the GemY company project and how smoothly the deployment went. Your expertise and experience with this particular product could be invaluable for my current project as I am operating under tight deadlines and lack the necessary time for complete product training. I have downloaded and reviewed the basic design guidelines and product description, but I am still unsure about potential risks, limitations, or additional integration steps that may be necessary. I feel a bit overwhelmed and anxious about the situation, as the project deadline was committed without considering the learning curve for a newproduct like this. I would greatly appreciate if you could share your insights and experiences regarding this product. If possible, could we arrange a call to discuss my questions and concerns in more detail? Your guidance

could be crucial in ensuring the success of this project. Thank you in advance for your help. I am looking forward to hearing from you soon. Best Regards, Andrew.

Email: Hi Lisa, I hope you're well. I would like to request your assistance in gathering some crucial data for our project. Our team's performance and cost efficiency need to be evaluated due to a sudden increase in the hourly rate this month. I believe it is important for us to be aware of our financial status, particularly when it comes to our project's budget.

Can you please download and provide me with the actual cost information from our ERP system? The default table might not contain all the necessary data, so I kindly request you to adjust the ERP extract parameters to ensure the downloaded file includes each team member ID, the hourly cost for the specific month, and the number of hours booked.

Moreover, I would really appreciate it if you could prepare a separate summary table comparing our planned and actual spending. Specifically, this should encompass the first five tasks which are already completed, depicting both the planned and spent cost and days for each. The remaining five tasks, which are yet to be completed, should only include the planned data.

I'm feeling quite anxious about our current financial situation, especially considering the recent rate increase. This analysis is imperative to help us strategically plan the remaining half of this project. I trust in your ability to convey this information accurately and promptly.

Thank you for your understanding and prompt attention to this matter. Best Regards.

Email: Dear Heather and Mike, I hope this message finds you well. We are currently working on a bid for a consulting data analytics project for Alltime, and your expertise in this field is greatly needed.

In light of the recent experience with underestimating project costs, I have decided to use an analogous estimating method for this project. The idea is to use activities from similar past projects to approximate the workload of the current project activities. For this, I have created a spreadsheet in our shared folder, which includes columns for workload estimation, required tools, and notably, a "Past Experience" column.

The "Past Experience" column is where I would like your assistance. Could you please populate it with the names of similar projects we have previously undertaken, their major risks, and the resulting workload? It is critical that these entries are descriptive enough to justify the use of that past record for estimating tasks in the new project.

I must admit, the results from the last project have left me feeling a bit uneasy. I believe thorough planning and precision in our estimations is crucial to avoid a similar situation and to meet both our cost targets and team bonuses. I am confident that with your help, we will be able to produce a more accurate estimation for the Alltime project.

Thank you for your assistance in this matter, it is greatly appreciated. Best regards, Andrew.

Data availability

Data will be made available on request.

References

Argaman, O., 2010. Linguistic markers and emotional intensity. J. Psycholinguist. Res. 39, 89–99.

Avelino, F., Hielscher, S., Strumińska-Kutra, M., de Geus, T., Widdel, L., Wittmayer, J., Dańkowska, A., Dembek, A., Fraaije, M., Heidary, J., et al., 2023. Power to, over and with: exploring power dynamics in social innovations in energy transitions across Europe. Environ. Innov. Soc. Transit. 48, 100758.

Boroditsky, L., 2011. How language shapes thought. Sci. Am. 304, https://doi.org/10.1038/scientificamerican0211-62

Braun, N., Goudbeek, M., Krahmer, E., 2020. Emotional words–the relationship of self- and other-annotation of affect. In: Proceedings of the Annual Meeting of the Cognitive Science Society, 42 (0).

Bringas Colmenarejo, A., Nannini, L., Rieger, A., Scott, K.M., Zhao, X., Patro, G.K., Kasneci, G., Kinder-Kurlanda, K., 2022. Fairness in agreement with European values: an interdisciplinary perspective on AI regulation. In: Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society, pp. 107–118.

Brundin, E., Patzelt, H., Shepherd, D.A., 2008. Managers' emotional displays and employees' willingness to act entrepreneurially. J. Bus. Ventur. 23, 221–243.

Buchner, Axel and Erdfelder, Edgar and Faul, Franz and Lang, Albert-Georg , 2007. G*Power. https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower.

Buechel, S., Hahn, U., EmoBank: studying the impact of annotation perspective and representation format on dimensional emotion analysis, arXiv preprint arXiv:2205.01996, 2022.

Byron, K., 2008. Carrying too heavy a load? The communication and miscommunication of emotion by email. https://doi.org/10.5465/amr.2008.31193163

Cambon, A., Hecht, B., Edelman, B., Ngwe, D., Jaffe, S., Heger, A., Vorvoreanu, M., Peng, S., Hofman, J., Farach, A., et al., 2023. Early LLM-based tools for enterprise information workers likely provide meaningful boosts to productivity. White Paper, Microsoft 4.

Carrasco-Farre, C., Large language models are as persuasive as humans, but why? about the cognitive effort and moral-emotional language of LLM arguments, arXiv preprint arXiv:2404.09329, 2024.

Clarke, V., Braun, V., 2021. Thematic analysis: a practical guide. https://doi.org/10.1177/1035719X211058251

Cohen, W.W., Carvalho, V.R., Mitchell, T.M., 2004. Learning to classify email into "speech acts". In: Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 309–316.

Constantinides, M., Verma, H., Sadeghian, S., El Ali, A., 2025. The future of work is blended, not hybrid. In: Proceedings of the 4th Annual Symposium on Human-Computer Interaction for Work, pp. 1–13.

Das, N., Shankar, S., Dash, B., Mohan, J., Pandey, M., Rautaray, S.S., 2021. Productivity profiling of organizations based upon communication interpretation and analysis. In: Proceedings of the International Conference on Information Systems and Computer Networks (ISCON). IEEE, pp. 1–6.

Dell'Acqua, F., McFowland, E., Mollick, E.R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., Lakhani, K.R., 2023. Navigating the jagged technological frontier: field experimental evidence of the effects of AI on knowledge worker productivity and quality. SSRN Electron. J. https://doi.org/10.2139/ssrn.4573321

Devillers, L., Vidrascu, L., Lamel, L., 2005. Challenges in real-life emotion annotation and machine learning based detection. Neural Netw. 18, 407–422.

DeWall, C.N., Baumeister, R.F., Mead, N.L., Vohs, K.D., 2011. How leaders self-regulate their task performance: evidence that power promotes diligence, depletion, and disdain. J. Pers. Soc. Psychol. 100, 47.

Dhillon, P.S., Molaei, S., Li, J., Golub, M., Zheng, S., Robert, L.P., 2024. Shaping human-AI collaboration: varied scaffolding levels in co-writing with language models. In: Proceedings of the CHI Conference on Human Factors in Computing Systems, pp. 1–18.

Dobel, C., Diesendruck, G., Bölte, J., 2007. How writing system and age influence spatial representations of actions: a developmental, cross-linguistic study: research report. Psychol. Sci. 18, https://doi.org/10.1111/j.1467-9280.2007.01926.x

Dong, W., Ehrlich, K., Macy, M.M., Muller, M., 2016. Embracing cultural diversity: online social ties in distributed workgroups. In: Proceedings of the ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW), pp. 274–287.

Farid Adilazuarda, M., Mukherjee, S., Lavania, P., Singh, S., Fikri Aji, A., O'Neill, J., Modi, A., Choudhury, M., 2024. Towards measuring and modeling "culture" in LLMs: a survey. arXiv e-prints, arXiv–2403. https://doi.org/10.18653/v1/2024.emnlp-main.882

Gandhi, V., Gandhi, S., Prompt sentiment: the catalyst for LLM change, arXiv preprint arXiv:2503.13510, 2025.

Ghosh, A., 2011. Power distance in organizational contexts-a review of collectivist cultures. Indian J. Ind. Relat. 89–101.

Grgić-Hlača, N., Castelluccia, C., Gummadi, K.P., 2022. Taking advice from (Dis) similar machines: the impact of human-machine similarity on machine-assisted decision-making. In: Proceedings of the AAAI Conference on Human Computation and Crowdsourcing, pp. 74–88.

Gunes, H., Pantic, M., 2010. Automatic, dimensional and continuous emotion recognition. Int. J. Synth. Emot. 1, 68–99.

Hall, E.T., 1976. Beyond culture. New York, 1981. ISBN-10 0385087470.

Hancock, J.T., Landrigan, C., Silver, C., 2007. Expressing emotion in text-based communication. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 929–932.

He, Y., Zhao, C., Hinds, P., 2010. Understanding information sharing from a cross-cultural perspective, in: CHI'10 Extended Abstracts on Human Factors in Computing Systems (CHI), pp. 3823–3828. https://doi.org/10.1145/1753846.1754063

Hofstede, G., 2011. Dimensionalizing cultures: the Hofstede model in context. Online Read. Psychol. Cult. 2, 8.

Hohenstein, J., Kizilcec, R.F., DiFranzo, D., Aghajari, Z., Mieczkowski, H., Levy, K., Naaman, M., Hancock, J., Jung, M.F., 2023a. Artificial intelligence in communication impacts language and social relationships. Sci. Rep. 13, https://doi.org/10.1038/s41598-023-30938-9

Hohenstein, J., Kizilcec, R.F., DiFranzo, D., Aghajari, Z., Mieczkowski, H., Levy, K., Naaman, M., Hancock, J., Jung, M.F., 2023b. Artificial intelligence in communication impacts language and social relationships. Sci. Rep. 13, 5487.

Holtbrügge, D., Weldon, A., Rogers, H., 2013. Cultural determinants of email communication styles. Int. J. Cross Cult. Manag. 13, 89–110.

Hoque, M.N., Mashiat, T., Ghai, B., Shelton, C.D., Chevalier, F., Kraus, K., Elmqvist, N., 2024. The hallmark effect: supporting provenance and transparent use of large language models in writing with interactive visualization. In: Proceedings of the CHI Conference on Human Factors in Computing Systems, pp. 1–15.

Hsu, Y.-C., Verma, H., Mauri, A., Nourbakhsh, I., Bozzon, A., et al., 2022. Empowering local communities using artificial intelligence. Patterns 3.

Inel, O., Draws, T., Aroyo, L., 2023. Collect, measure, repeat: reliability factors for responsible AI data collection. In: Proceedings of the AAAI Conference on Human Computation and Crowdsourcing, pp. 51–64.

Jackson, J.C., Watts, J., Henry, T.R., List, J.-M., Forkel, R., Mucha, P.J., Greenhill, S.J., Gray, R.D., Lindquist, K.A., 2019. Emotion semantics show both cultural variation and universal structure. Science 366, 1517–1522.

Jakesch, M., Hancock, J.T., Naaman, M., 2023. Human heuristics for AI-generated language are flawed. Proc. Natl. Acad. Sci. U.S.A. 120, https://doi.org/10.1073/pnas.2208839120

Jovic, M., Mnasri, S., 2024. Evaluating AI-generated emails: a comparative efficiency analysis. World J. Engl. Lang. 14.

Kadoma, K., Quere le, M.A., Fu, J., Munsch, C., Metaxa, D., Naaman, M., The role of inclusion, control, and ownership in workplace AI-mediated communication, arXiv preprint arXiv:2309.11599, 2023.

Kannan, A., Kurach, K., Ravi, S., Kaufmann, T., Tomkins, A., Miklos, B., Corrado, G., Lukacs, L., Ganea, M., Young, P., et al., 2016. Smart reply: automated response suggestion for email. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), pp. 955–964.

Khatri, N., 2009. Consequences of power distance orientation in organisations. Vision 13, 1–9.

Kiuru, N., Spinath, B., Clem, A.-L., Eklund, K., Ahonen, T., Hirvonen, R., 2020. The dynamics of motivation, emotion, and task performance in simulated achievement situations. Learn. Individ. Differ. 80, 101873.

Kokkalis, N., Köhn, T., Pfeiffer, C., Chornyi, D., Bernstein, M.S., Klemmer, S.R., 2013. EmailValet: managing email overload through private, accountable crowdsourcing. In: Proceedings of the 2013 Conference on Computer Supported Cooperative Work, pp. 1291–1300.

Kolisko, S., Anderson, C.J., 2023. Exploring social biases of large language models in a college artificial intelligence course. https://doi.org/10.1609/aaai.v37i13.26879

Kort, B., Reilly, R., Picard, R.W., 2001. An affective model of interplay between emotions and learning: reengineering educational pedagogy-building a learning companion. https://doi.org/10.1109/ICALT.2001.943850

Lampert, A., Dale, R., Paris, C., 2008. Requests and commitments in email are more complex than you think: eight reasons to be cautious. In: Australasian Language Technology Association Workshop. Australasian Language Technology Association, pp. 64–72.

Lee, M., Liang, P., Yang, Q., 2022. CoAuthor: designing a human-AI collaborative writing dataset for exploring language model capabilities. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, pp. 1–19.

Li, C., Chen, M., Wang, J., Sitaram, S., Xie, X., 2024. CultureLLM: incorporating cultural differences into large language models. Adv. Neural Inf. Process. Syst. 37, 84799–84838.

Manam, V.K., Quinn, A., 2018. Wingit: efficient refinement of unclear task instructions. In: Proceedings of the AAAI Conference on Human Computation and Crowdsourcing, pp. 108–116.

McKinsey, 2012. The social economy: unlocking value and productivity through social technologies. https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/the-social-economy.

Meng, J., Dai, Y.N., 2021. Emotional support from AI chatbots: should a supportive partner self-disclose or not? J. Comput.-Mediat. Commun. 26, 207–222. https://doi.org/10.1093/jcmc/zmab005

Meyer, E., 2014. The culture map - decoding how people think, lead and get things done across culture. https://doi.org/10.31128/ajgp-09-24-7410

Mohammad, S., 2018. Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 english words. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 174–184.

Mohammad, S.M., Turney, P.D., 2013. Crowdsourcing a word–emotion association lexicon. Comput. Intell. 29, 436–465.

Nugroho, S., Sitorus, A.T., Habibi, M., Wihardjo, E., Iswahyudi, M.S., 2023. The role of ChatGPT in improving the efficiency of business communication in management science. J. Minfo Polgan 12, 1482–1491.

OpenAI, R, GPT-4 technical report, arXiv preprint arXiv:2303.08774, 2023. View in Article 2, 1.

Pekerti, A.A., Thomas, D.C., 2003. Communication in intercultural interaction: an empirical investigation of idiocentric and sociocentric communication styles. J. Cross-Cult. Psychol. 34, 139–154.

Peng, S., Kalliamvakou, E., Cihon, P., Demirer, M., The impact of AI on developer productivity: evidence from github copilot, arXiv preprint arXiv:2302.06590, 2023.

Pervaiz, S., Guohao, L., Qi, H., 2024. Task-crafting: how power distance shapes the influence of goal-setting participation. Curr. Psychol. 1–13.

Plutchik, R., 1980. A general psychoevolutionary theory of emotion. In: Theories of Emotion. Elsevier, pp. 3–33. https://doi.org/10.1016/B978-0-12-558701-3.50007-7

Plutchik, R., 1982. A psychoevolutionary theory of emotions. Soc. Sci. Inf. 21, https://doi.org/10.1177/053901882021004003

Plutchik, R., 1984. Emotions: a general psychoevolutionary theory. Approaches to emotion 1984, 2–4.

Plutchik, R., 2004. Emotions in the practice of psychotherapy: clinical implications of affect theories. https://psycnet.apa.org/doi/10.1037/10366-000.

Project Management Institute, 2021. The Standard for Project Management and a Guide to the Project Management Body of Knowledge (PMBOK), seventh ed. ISBN-10 1628256648.

Renze, M., 2024. The effect of sampling temperature on problem solving in large language models. In: Findings of the Association for Computational Linguistics: EMNLP 2024, pp. 7346–7356.

Russell, J.A., 1991. Culture and the categorization of emotions. Psychol. Bull. 110, https://doi.org/10.1037/0033-2909.110.3.426

Salinas, A., Morstatter, F., The butterfly effect of altering prompts: how small changes and jailbreaks affect large language model performance, arXiv preprint arXiv:2401.03729, 2024.

Santiago Jr, C.S., Embang, S.I., Conlu, M.T.N., Acanto, R.B., Lausa, S.M., Ambojia, K.W.P., Laput y, E., Aperocho, M.D.B., Malabag, B.A., Balilo Jr, B.B., et al., 2023. Utilization of writing assistance tools in research in selected higher learning institutions in the Philippines: a text mining analysis. Int. J. Learn. Teach. Educ. Res. 22, 259–284.

Sappelli, M., Pasi, G., Verberne, S., de Boer, M., Kraaij, W., 2016. Assessing e-mail intent and tasks in e-mail messages. Inf. Sci. 358-359, 1–17. https://doi.org/10.1016/j.ins.2016.03.002

Sekhon, H., Ennew, C., Kharouf, H., Devlin, J., 2014. Trustworthiness and trust: influences and implications. J. Mark. Manag. 30, 409–430.

Semin, G.R., Görts, C.A., Nandram, S., Semin-Goossens, A., 2002. Cultural perspectives on the linguistic representation of emotion and emotion events. Cogn. Emot. 16, 11–28.

Tarakci, M., Greer, L.L., Groenen, P.J.F., 2016. When does power disparity help or hurt group performance? J. Appl. Psychol. 101, 415.

Tirado-Olivares, S., Navío-Inglés, M., O'Connor-Jiménez, P., Cózar-Gutiérrez, R., 2023. From human to machine: investigating the effectiveness of the conversational AI ChatGPT in historical thinking. Educ. Sci. 13, 803.

Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., et al., LLaMA: open and efficient foundation language models, arXiv preprint arXiv:2302.13971, 2023.

Treiman, L.S., Ho, C.-J., Kool, W., 2023. Humans forgo reward to instill fairness into AI. In: Proceedings of the AAAI Conference on Human Computation and Crowdsourcing, pp. 152–162.

Venkit, P.N., Gautam, S., Panchanadikar, R., Huang, T.-H., Wilson, S., Nationality bias in text generation, arXiv preprint arXiv:2302.02463, 2023.

Weber, F., Wambsganss, T., Neshaei, S.P., Soellner, M., 2024. LegalWriter: an intelligent writing support system for structured and persuasive legal case writing for novice law students. In: Proceedings of the CHI Conference on Human Factors in Computing Systems, pp. 1–23.

Wiggins, B.E., 2012. Toward a model for intercultural communication in simulations. Simul. Gaming 43, https://doi.org/10.1177/1046878111414486

Zakaria, N., 2017. Emergent patterns of switching behaviors and intercultural communication styles of global virtual teams during distributed decision making. J. Int. Manag. 23, 350–366.