Computers in Human Behavior: Artificial Humans 9 (2026) 100362

Contents lists available at ScienceDirect

Computers in Human Behavior: Artificial Humans

journal homepage: www.journals.elsevier.com/computers-in-human-behavior-artificial-humans

Research paper

Mind the style: Impact of communication style on human-chatbot interaction

Erik Derner [a,b], Dalibor Kučera [c], Aditya Gulati [a], Robert A. Bagheri [d,*], Nuria Oliver [a]

[a] ELLIS Alicante, Alicante, Spain
[b] Czech Technical University in Prague, Prague, Czech Republic
[c] University of South Bohemia, České Budějovice, Czech Republic
[d] Utrecht University, Utrecht, The Netherlands

ARTICLE INFO

Keywords:
Conversational agents
Communication style
Generative AI

ABSTRACT

Conversational agents increasingly mediate everyday digital interactions, yet the effects of their communication style on user experience and task success remain insufficiently understood. Addressing this gap, we report a between-subject user study in which participants interacted with one of two versions of a chatbot called NAVI, which assisted them in an interactive map-based 2D navigation task. The two chatbot versions were designed to differ primarily in communication style: one used a friendly and supportive tone, while the other used a direct and task-focused tone. We also included a control condition where participants did not interact with a chatbot but received the step-by-step navigation instructions. The friendly chatbot significantly increased users' communication satisfaction and was associated with higher task success than the direct chatbot. However, participants in the control condition achieved the highest task success overall, suggesting that chatbot interaction may introduce overhead in tasks that can be completed effectively using straightforward instructions. We did not find significant evidence that gender moderated the effects of communication style, although exploratory gender-stratified analyses suggested patterns that warrant further investigation. Finally, we found limited evidence of global linguistic accommodation, with only selective feature-level alignment. These findings suggest that chatbot communication style influences users' perceptions of conversational agents and may improve performance relative to less supportive chatbot designs, but the overall value of chatbot interaction depends on the task context. The study highlights the need for task-sensitive, transparent and carefully evaluated communication-style choices in conversational-agent design.

1. Introduction

Conversational agents are increasingly used for task-oriented assistance, including in navigation, learning and decision-making scenarios (Chowa et al., 2025; Yi et al., 2025). Their effectiveness depends not only on informational accuracy but also on their communication style. Stylistic cues such as tone, politeness, warmth and brevity can shape users satisfaction, engagement and their judgment of the agent's competence, making communication style an important design variable in human–AI interaction (Cai et al., 2024; Ding & Najaf, 2024; Feine et al., 2019; Go & Sundar, 2019).

In this paper, we use the term communication style in an operational sense. We do not treat "friendly" and "direct" as universal interpersonal types, nor as stable personalities of the conversational agent. Rather, they refer to two prompt-engineered modes of expressing the same navigation guidance. The friendly style operationalizes warmth, supportiveness, politeness and encouragement, whereas the direct style operationalizes brevity, task focus and minimal social framing. This distinction is grounded in work on politeness and relational communication (Brown & Levinson, 1987), social responses to media (Reeves & Nass, 1996), and socio-cognitive models of warmth and competence (Fiske et al., 2002). It is also practically relevant because contemporary Large Language Model (LLM) based chatbots can be instructed to vary style while keeping their task goal constant.

Social Response Theory proposes that people apply social norms to technologies that display human-like cues, even when they know that the system is artificial (Nass & Moon, 2000; Reeves & Nass, 1996). Thus, linguistic style functions as a social signal that shapes the users' affective reactions and expectations. Communication Accommodation Theory complements this view by explaining how interlocutors adjust their linguistic behavior (*e.g.,* lexical choices, pronouns, affective expressions, directness or temporal/numerical references) to manage affiliation, reduce social distance or express identity (Giles et al., 2023; Giles & Ogay, 2013). Expectancy-based models additionally suggest that communication style matters because it can confirm or violate the users' assumptions about how an interaction should unfold (Burgoon

* Corresponding author.
E-mail address: a.bagheri@uu.nl (R.A. Bagheri).

https://doi.org/10.1016/j.chbah.2026.100362
Received 3 February 2026; Received in revised form 8 July 2026; Accepted 15 July 2026
Available online 31 July 2026
2949-8821/© 2026 Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

& Hubbard, 2005). These frameworks imply that stylistic variation in conversational agents may affect how interactions are perceived and evaluated, shaping impressions of warmth, competence, social presence, and also eliciting linguistic alignment or accommodation (Sterken & Kirkpatrick, 2024).

Furthermore, prior research suggests that the stylistic appropriateness of conversational agents is context-dependent: informal or friendly styles can increase perceived social presence, yet may also be judged as unsuitable in highly goal-directed tasks (Liebrecht et al., 2020). Moreover, perceived "human-likeness" in conversational agents appears to rely less on warmth than on clarity, consistency and linguistic precision (Svenningsson & Faraon, 2019). While prior research has examined dimensions such as empathy, politeness and anthropomorphism in conversational agents (Bickmore & Picard, 2005), relatively few studies have systematically isolated *fixed stylistic tone* in a controlled, goal-oriented task where performance can be objectively measured while task content is held constant. Moreover, it remains unclear whether linguistic accommodation emerges in short, tightly constrained interactions, or whether user characteristics – such as gender or prior experience with chatbots – condition responses to communication style. Recent findings further suggest that user responses to stylistic variation depend on contextual appropriateness and the users' awareness of interacting with an artificial agent (Islind et al., 2023), underscoring the need for controlled experimental evidence.

In this paper, we focus on chatbots as one class of conversational agents. We investigate how a chatbot's communication style influence task performance, interaction behavior, and subjective user evaluations within a controlled map-based 2D navigation task. To this end, we design two stylistic variants of a chatbot `NAVI`: a friendly and supportive style ($\texttt{NAVI}_f$) or a direct, task-focused style ($\texttt{NAVI}_d$). These two chatbot conditions were compared with a control condition in which participants received static step-by-step instructions without a chatbot. In addition, we investigate whether users linguistically accommodate to the chatbot's communication style and whether such effects are moderated by individual differences, including prior experience with chatbots and gender. The design allows us to examine how stylistic tone relates to subjective satisfaction, task success and linguistic behavior while keeping the navigation route and task goal constant.

Our contributions are threefold: (1) we provide evidence on how fixed communication style relates to both relational (communication satisfaction) and instrumental (task success) in a measurable, controlled and goal-oriented task; (2) we examine whether users linguistically accommodate to the chatbot's style in short, task-focused exchanges; (3) we study whether gender and prior chatbot experience shape responses to style, with implications for the design of inclusive and socially attuned conversational agents.

The rest of the paper is structured as follows: we provide a summary of the most relevant related work and formulate the research questions that drive our research in Section 2. Our user study and its results are presented in Sections 3 and 4, respectively. We provide a discussion of our findings in Section 5, followed by limitations and future work in Section 6, and conclusion in Section 7.

2. Related work

2.1. Communication style in conversational agents

Research in HCI demonstrates that users respond socially to systems that employ human-like cues (Nass & Moon, 2000; Reeves & Nass, 1996). Conversational style, understood as the patterned use of tone, politeness, affective framing, directness and interactional support in dialogue, influences the user's perceptions of a conversational agent's personality, appropriateness and competence (Feine et al., 2019; Mairesse & Walker, 2011; Pralat et al., 2024). Friendly or warm styles tend to increase perceived social presence and relational engagement, whereas more direct or task-focused tones may enhance perceived clarity and efficiency while reducing perceived warmth (Bhattacharjee et al., 2025; Deng & Yan, 2025; Go & Sundar, 2019). However, stylistic effectiveness is not universal: informal or friendly styles can increase social presence but may be judged as inappropriate in highly instrumental contexts (Liebrecht et al., 2020), and perceived humanness in task-oriented chatbots may depend strongly on linguistic precision rather than emotional expressiveness (Svenningsson & Faraon, 2019).

This literature suggests that communication style should be treated as a design parameter whose effects depend on task demands and user expectations. From a socio-cognitive perspective, warmth and competence are two central dimensions of social judgment (Fiske et al., 2007). In human–AI interaction, warmth has been linked to affective trust and satisfaction, while competence supports performance-related trust (Kulms & Kopp, 2018; Lee & See, 2004). The broader quality of communication experience, including clarity, responsiveness and relational cues, offers an additional lens for understanding stylistic effects (Gray & Laidlaw, 2004). Thus, a friendly style may improve subjective communication experience without necessarily improving objective task performance, whereas a direct style may be efficient but less relationally satisfying. This distinction is central for the present study.

In addition, most prior work has examined adaptive, affective or multimodal agents (Gnewuch et al., 2018; Lee et al., 2011). Less attention has been given to fixed, non-adaptive styles in controlled tasks where objective outcomes can be measured and where the informational route is held constant (Diederich et al., 2022; Qin et al., 2025). This limits our understanding of whether stylistic tone alone affects the users' subjective evaluations of the interaction, and whether such effects extend to task performance. It also matters whether users know they are interacting with a chatbot rather than a human, because this awareness can shape expectations and evaluations of style (Islind et al., 2023). The present study focuses on a chatbot that was explicitly presented to participants as a conversational agent and manipulates its communication style.

2.2. Social alignment and linguistic accommodation

Communication Accommodation Theory (CAT) describes how individuals adjust lexical, syntactic and paralinguistic behavior to manage affiliation or distinction (Giles et al., 2023; Giles & Ogay, 2013). Evidence of linguistic alignment with artificial interlocutors has been observed in text-based and voice-based interactions, including mirroring of pronouns, affective expressions and syntactic structures. However, such accommodation is often selective and feature-specific rather than reflecting global convergence of style (Luger & Sellen, 2016). Expectancy-driven models similarly propose that communicative behavior is shaped by anticipated interactional norms, with alignment or compensation depending on whether the agent's style confirms or violates these expectations (Burgoon & Hubbard, 2005). Congruence between agent tone and user expectations can strengthen rapport and social presence, whereas mismatches can hinder cooperation (Feine et al., 2019; Nass et al., 1996; Ruijten et al., 2015).

The practical relevance of linguistic accommodation is that it helps distinguish two kinds of stylistic influence. A chatbot's style may affect how users evaluate the interaction, but it may or may not shape how users themselves communicate during the interaction. If users accommodate to an agent's style, stylistic choices can alter the interaction dynamics and not only post-task impressions. If accommodation is limited, style may operate mainly through perception, satisfaction and task engagement. Most existing evidence comes from longer, richer or multimodal interactions. It therefore remains unclear whether accommodation emerges in short, goal-oriented text exchanges where users focus primarily on solving a task.

2.3. Gender and relational orientation in HCI

Sociolinguistic and psychological research has reported average gender-related differences in communication goals, with women often described as more oriented toward relational and affiliative cues and men as more oriented toward instrumental aspects such as efficiency or control (Cross & Madson, 1997; Tannen, 1990). In technology use, prior work similarly suggests that women tend to be more sensitive to socially expressive interfaces, while men evaluate systems more through usefulness and ease of use (Gefen & Straub, 1997; Sobieraj & Krämer, 2020). Studies further show that women experience higher social presence and relational engagement when interacting with polite or friendly agents (Ruijten et al., 2015), suggesting that communication style may interact with gender to shape satisfaction and trust (Ruijten et al., 2015).

At the same time, gender-related differences in HCI are typically context-dependent, can be impacted by task design and prior experience (Nazareth et al., 2019), and may not appear as statistically reliable moderation effects in a given experiment. For this reason, the present study examines gender as a potential user characteristic that may condition responses to style, but we treat gender-stratified patterns cautiously and distinguish them from confirmed interaction effects.

2.4. Effects of prior chatbot experience on expectations and evaluations

Beyond immediate stylistic cues, prior experience with conversational agents may shape how users perceive and evaluate chatbot interaction. Users with extensive prior experience tend to develop more calibrated mental models of system capabilities, which can lead to higher task efficiency but also more critical assessments of system performance and language quality (Luger & Sellen, 2016). In contrast, less experienced users may attribute greater intelligence or intentionality to chatbots, resulting in higher initial satisfaction but also increased risk of expectation violation when system limitations become apparent (Rheu et al., 2024). These differences suggest that prior chatbot experience may condition how communication style and system behavior are perceived.

Empirical studies further indicate that prior experience influences how users evaluate conversational style, social presence and trustworthiness. Experienced users tend to prioritize efficiency, clarity and goal alignment over relational or anthropomorphic cues, whereas novice users may respond more positively to friendly or supportive language that scaffolds the interaction and reduces uncertainty (Islind et al., 2023). Experience-related differences can also affect users' willingness to adapt their own language, with experienced users showing less linguistic accommodation and greater reliance on concise, task-oriented commands (Bell, 2003; Zellou & Holliday, 2024). Finally, longitudinal evidence suggests that repeated exposure and regular interactions with conversational agents influence perceptions of trust, dialogue quality, anthropomorphism and interaction behavior (Araujo & Bol, 2024).

2.5. Research questions

Guided by Social Response Theory, Communication Accommodation Theory and socio-cognitive models of warmth and competence, the present study examines how fixed communication style affects user performance, experience and linguistic behavior when other interactional factors are held constant. We investigate the following research questions:

RQ1 - communication style and user outcomes. How does a chatbot's communication style (friendly vs. direct) influence task performance and satisfaction in a goal-oriented map-based 2D navigation task? This question targets both relational and instrumental effects of stylistic tone.

RQ2 – moderation by user characteristics. Do user characteristics, such as gender and previous chatbot experience, moderate the influence of communication style on task and experience measures? Given prior work on relational versus instrumental orientations and chatbot expectations, we examine whether sensitivity to stylistic tone differs across user groups.

RQ3 – linguistic accommodation. Do users linguistically adapt to the chatbot's communication style during short, task-focused interactions? By analyzing lexical and stylistic features in user messages, we test whether style affects the users' own language or primarily their evaluations of the interaction.

We address these three research questions by means of a between-subject user study designed to evaluate communication style as both a social cue and a functional design parameter, and to assess its relationship with user characteristics and linguistic behavior.

3. User study

3.1. Task: Map-based 2D navigation as a testbed for communication style

Navigation tasks offer a structured environment for examining how users integrate external guidance into goal-oriented action. Unlike open-ended conversational settings, map-based 2D navigation constrains both the problem and the interaction space, thereby allowing stylistic tone to be manipulated independently of content and task structure. Prior work suggests that navigation performance is sensitive to cue structure, time pressure, and the degree of required initiative, *e.g.,* distinguishing stepwise instruction-following from dialogue-based exploration (Brunyé et al., 2017; Dahmani et al., 2023).

A controlled navigation task therefore offers a tractable experimental paradigm to study communication style as the primary independent variable, while holding constant task complexity, interaction modality, and informational content. To this end, we designed a map-based 2D navigation task in which participants were required to locate a predefined target destination (the embassy) using concise, content-equivalent textual guidance (See Fig. 3).

In addition, we included a control condition in which participants received only step-by-step navigation instructions without interacting with a chatbot. This control condition served two purposes: it provided a baseline level of task performance in the absence of conversational interaction; and it allowed us to confirm that the navigation task itself did not systematically provide an advantage to particular user profiles. Note that the control condition was not intended to serve as a stylistic baseline, but rather as a validation of the task environment. Thus, comparisons between the two chatbot conditions allow for a focused assessment of the impact of communication style while holding all other task characteristics constant.

3.2. Design

We used a between-subjects experimental design, depicted in Fig. 1. The study consisted of two parts: (1) a navigation task and (2) a post-task survey that collected subjective feedback regarding communication quality, satisfaction, user experience and preferences for chatbot versus human assistance. Measures are described in detail in 3.7, and full item wordings are provided in Appendix B.

3.2.1. Map-based 2D navigation task

Participants were instructed to imagine arriving at the train station of a capital city where they had never been before. They realized that they did not have their phone and documents with them, so they needed to urgently reach their country's embassy. Their task was to figure out where the embassy was located by navigating from the train station to the embassy on a 2D map.

The interface displayed a static city map with streets, buildings and a large lake, together with a legend and compass rose indicating

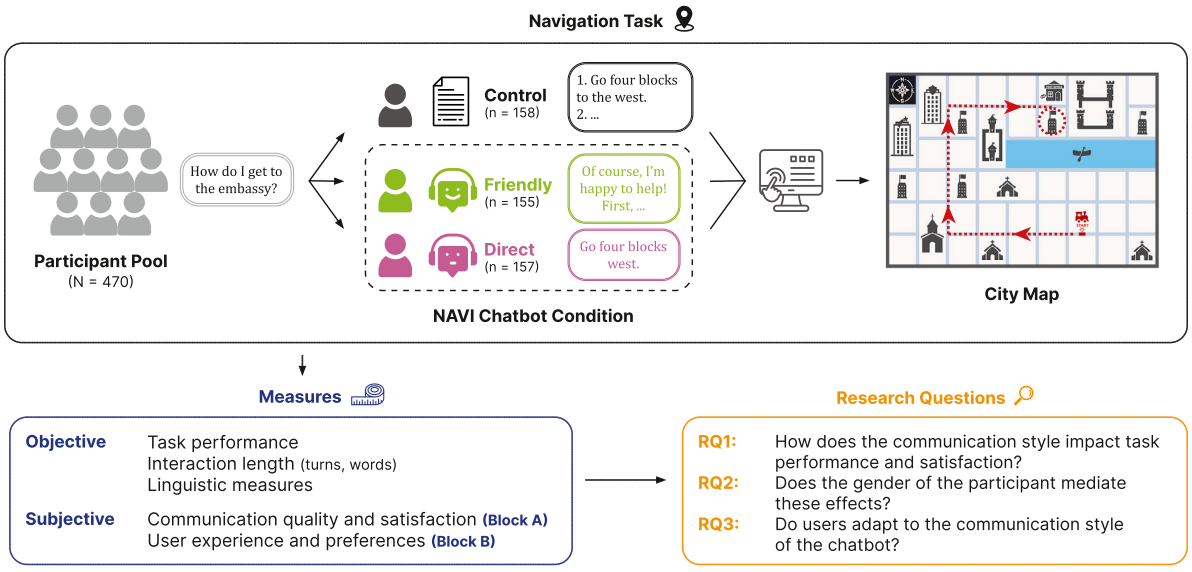


Fig. 1. Overview of the study design. Participants completed a fictional map-based 2D navigation task to locate an embassy on a map (top). They followed static instructions (**Control**) or they interacted with one of the two versions of the chatbot NAVI, differing only in the communication style: **Friendly** (supportive, warm) or **Direct** (concise, task-focused). Both objective and subjective measures (bottom-left) were collected to address three research questions (bottom-right).

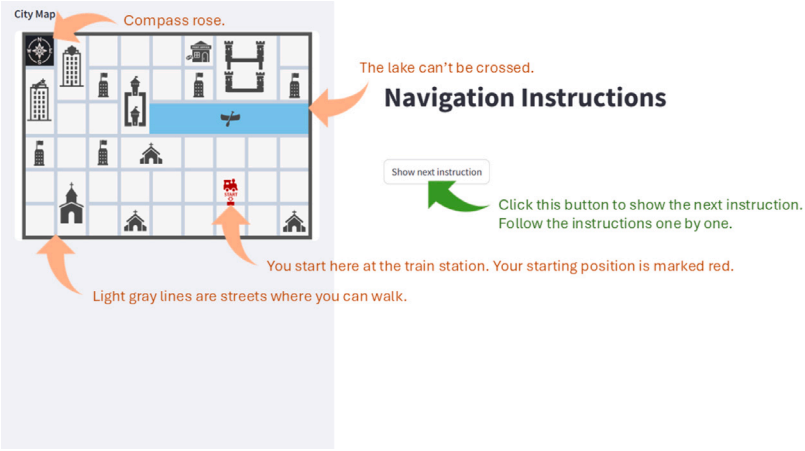


Fig. 2. Explanations of the user interface presented to the participants in the control condition, including the map legend.

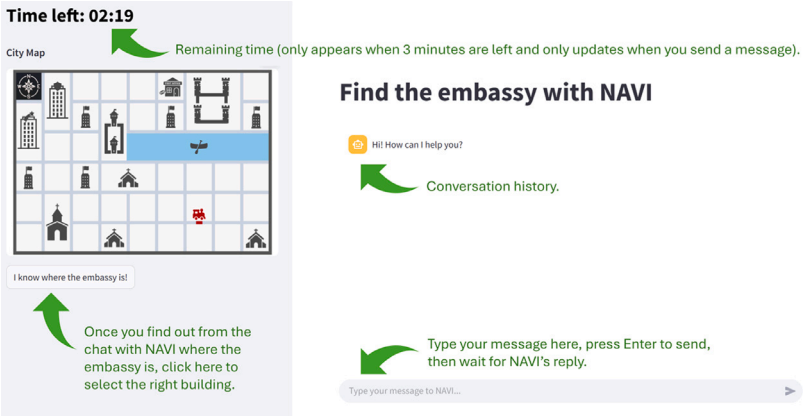


Fig. 3. Explanations of the user interface used by the participants to interact with NAVI.

cardinal directions. Streets were represented by light gray lines that could be used for walking, the lake could not be crossed and the starting position at the train station was marked with a red square. A detailed explanation of the interface provided participants with all necessary information about the map layout, clickable elements and interaction flow, as can be seen in Fig. 2.

Control condition (no chatbot). In the control condition, participants received step-by-step written instructions guiding them from the train station to the embassy. The full set of instructions is provided in Appendix C. The interface showed a static 2D city map on the left and a panel with navigation instructions on the right, as depicted in Fig. 2. Participants clicked a button to reveal the next instruction in sequence and were asked to follow these instructions on the map until they believed they had identified the embassy. The layout and visual elements were matched as closely as possible to the chatbot interface to ensure that the main difference between conditions was the presence of a conversational agent rather than differences in visual design or information structure.

Chatbot condition (NAVI). In the chatbot condition, participants were introduced to NAVI, a ChatGPT-based conversational agent with local knowledge that could help them find the embassy. NAVIwas embedded in a custom web interface that combined a chat window on the right with the same static 2D city map on the left, as illustrated in Fig. 3. As shown in the Figure, NAVI's icon was a robot to indicate that participants were interacting with a chatbot. Participants were told that they could ask NAVIfor directions and clarifications in order to locate the embassy.

Upon entering the chatbot branch of the study, participants were randomly assigned to one of two communication-style variants of NAVI:

- Friendly NAVI(NAVI_f), which was configured to use a warm, polite and supportive communication style. It offered encouragement, reiterated instructions when participants appeared confused, and would occasionally include brief contextual remarks (*e.g.,* about nearby landmarks) alongside the same route guidance. These remarks were intended to foster rapport and social presence rather than provide additional task-relevant navigation information.
- Direct NAVI(NAVI_d), which provided brief, task-focused responses that were efficient and to the point. This version avoided social niceties, did not offer encouragement and minimized contextual or relational commentary. The goal was to maximize clarity and efficiency, emphasizing instrumental communication over social engagement while conveying the same core navigation guidance.

The friendly communication operationalized friendliness through a broader set of conversational behaviors commonly associated with warm, rapport-oriented interactions, including a warmer tone and occasional local tips and contextual remarks while keeping the core navigation content consistent across conditions. Therefore, the manipulation reflects differences in overall communication style rather than differences in linguistic form alone.

Participants were not informed which version of NAVIthey were interacting with, and both versions followed the same underlying navigation route from the train station to the embassy. In all chatbot conditions, NAVI's behavior was controlled by style-specific system prompts, which specified communicative goals, politeness level, affective tone and constraints on content. The exact wording of the system prompts for NAVI_f and NAVI_d is provided in Appendix A. To keep the manipulation focused on communication style, we constrained the chatbot to reveal a fixed route step by step. The sequence of directions and landmarks was held constant across both styles, and NAVIwas instructed to avoid giving away the answer directly, for example, by naming the embassy building number.

Because the underlying model is generative, individual responses could still exhibit some stochastic variability. To verify that the stylistic manipulation was stable and implemented as intended, we later profiled NAVI's utterances using the Linguistic Inquiry and Word Count (LIWC) dictionary and complementary embedding-based measures (see Section 3.7). These analyses confirmed that the friendly and direct versions displayed distinct and consistent patterns in affective and relational categories, while within-style variation remained low relative to between-style differences.

Participants interacted with NAVIvia the chat interface to obtain directions from the starting point (the train station) to the target (embassy). The interface displayed a timer that became visible only when the remaining time dropped below three minutes, and this timer updated whenever participants sent a message. In the instructions, participants were told that, on average, six to eight conversational turns were usually needed to locate the embassy. To discourage attempts to solve the task without meaningful interaction, a minimum of three user–NAVIturns were enforced before participants could submit their final answer with the location of the embassy.

In both the control and chatbot conditions, participants indicated that they had reached a decision by clicking a button located below the map on the left-hand side of the interface with the label *"I know where the embassy is"*. Once they clicked this button, numbers from 1 to 15 appeared over all buildings on the map. Participants were then asked to select the number corresponding to the building they believed to be the embassy. Feedback about correctness was not provided at any point to avoid learning or feedback effects.

The navigation task had a maximum duration of eight minutes. If a participant had not submitted an answer by the end of this period, the task was terminated and the participant was automatically advanced to the post-task survey. The eight-minute limit was chosen based on pilot data, which indicated that this duration was sufficient for participants to complete the task under normal circumstances while still imposing a realistic sense of time pressure.

3.2.2. Post-task survey

After completing the navigation task and selecting the embassy building, participants proceeded to a post-task survey. The survey captured subjective evaluations of the interaction and provided additional context for interpreting objective task measures.

In the chatbot conditions, participants completed a questionnaire with randomized block order and embedded attention checks. The survey contained questions about perceived communication quality and satisfaction, perceived usefulness of NAVIfor accomplishing the task, and preferences for interacting with a chatbot versus a human in similar situations. It also captured chatbot use frequency to contextualize individual differences in user experience. The complete set of survey items, including the wording and response scales, is provided in Appendix B, and all derived measures are described in Section 3.7.

In the control condition, participants completed a shorter post-task survey that did not include communication quality items, since they had not interacted with a conversational agent. Instead, they evaluated their satisfaction with the written instructions, their preference for receiving directions from a human versus written instructions, their preference for using a chatbot instead of static instructions and their chatbot use frequency. The corresponding items for the control condition are listed in Appendix B (control survey section), again with measures described in Section 3.7.

3.3. Implementation

The study was implemented using Streamlit, an open-source Python framework for interactive web applications. Streamlit allowed us to integrate instructions, the map interface, the chat window and the post-task survey within a single application, which reduced context switching for participants and simplified logging of interaction data.

The application was hosted on Microsoft Azure, which provided scalable and reliable access for Prolific participants. Azure storage was used to securely log each participant's progression through the study, including navigation behavior, chat transcripts and survey responses. The chatbot was powered by the gpt-4.1-2025-04-14 model via the OpenAI API, which served as the backbone for generating NAVI's responses. The study was integrated with Prolific's recruitment and screening tools, enabling automatic enforcement of inclusion criteria and controlled access to the study link.

Participants were randomly assigned by the web application to one of the three experimental conditions (control, friendly NAVI, direct NAVI). Within the chatbot branch, assignment to friendly versus direct NAVIwas based on simple random allocation with equal target probability. To maintain approximately equal numbers of women and men in each condition, we additionally monitored recruitment through Prolific's prescreening and quota tools and adjusted ongoing recruitment so that gender distributions remained balanced across conditions.

3.4. Pilot studies

To refine the study design and ensure a smooth user experience, we conducted three iterative pilot studies on Prolific, each with 10 participants. After each pilot, we examined navigation performance, chat transcripts and survey responses, and adjusted the interface and protocol accordingly. Several key design choices were informed by these pilots:

- We introduced a minimum number of required conversational turns, because early pilots showed that some participants tried to guess the embassy location without interacting meaningfully with NAVI.
- We refined the time estimates and compensation to align with the actual time needed to complete the task and survey, and we updated the information provided to participants so that expectations about study duration were realistic.
- We clarified and streamlined the instructions, including the narrative scenario, explanation of the interface and examples of how to ask NAVIfor help.
- We adjusted the color scheme and contrast of the map to improve readability and reduce visual ambiguity.

While the map layout (landmark positions) and the style-specific system prompts were initially developed through exploratory prototyping, both were fixed prior to the pilot phase and remained unchanged across all pilot and main study sessions. The final version of the instructions shown to participants is included in Appendix C.

3.5. Sample size and power considerations

Before data collection, we conducted *a priori* power considerations to determine the sample size. Prior work on stylistic manipulations in conversational agents typically reports effects in the small-to-medium range on relational outcomes such as trust and satisfaction, with Cohen's d values often between 0.25 and 0.40. Assuming a two-sided α of 0.05 and desired power of 0.80, a medium effect ($d \approx 0.50$) can be detected with about 64 participants per group, whereas detecting a lower-medium effect ($d \approx 0.30$) requires larger groups of around n = 170–180 participants per condition. For correlations between linguistic indicators and subjective evaluations, similar calculations show that medium correlations ($r \approx 0.30$) require approximately n = 80–90 participants, while smaller correlations ($r \approx 0.20$) require about n = 190–200 participants. Anticipating effects in the small-to-medium range and aiming to support both between-condition comparisons and exploratory analyses of linguistic alignment and gender moderation, we targeted a total sample of approximately 450 to 500 participants distributed across the three conditions. The final sample of 470 participants (see Section 3.8) thus provides adequate sensitivity for detecting stylistic effects of theoretical interest and sufficient stability for the text-based analyses.

3.6. Ethical approval

The full study protocol, including recruitment, consent materials, data handling and debriefing procedures, was reviewed and approved by the ethics committee of the Faculty of Education, University of South Bohemia (approval reference number: EK049/2025, dated April 28, 2025). The committee concluded that the project complied with institutional regulations and international standards for research with human participants. All participants provided informed consent prior to beginning the study and were compensated in line with Prolific guidelines and the estimated study duration (see Fig. 4).

3.7. Measures

We collected objective indicators of task performance and interaction behavior, self-reported evaluations of the interaction, and linguistic measures capturing structural, stylistic and affective aspects of the interaction. This combination of measures was designed to address our three research questions by jointly characterizing instrumental outcomes (task success and persistence), relational outcomes (satisfaction) and linguistic accommodation.

3.7.1. Objective measures

We collected the following objective measures:

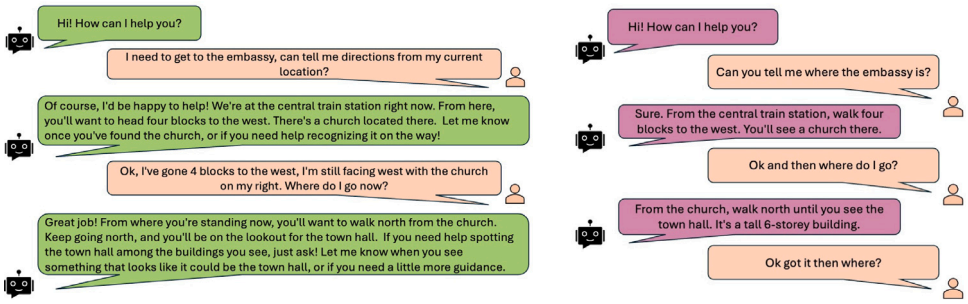
1. Task performance, defined as a binary indicator of whether participants correctly identified the embassy building at the end of the navigation task. This measure directly captures the primary instrumental outcome of the guidance and provides a simple, interpretable criterion for comparing conditions.

2. Interaction behavior, measured by the number of conversational turns between the participant and NAVI, the number of explicit help requests and the total time spent on the navigation task. Because reaction-time data collected in uncontrolled online environments are often noisy, we treated time as a secondary contextual measure rather than a primary efficiency metric. The main behavioral indicators were therefore the extent of conversational engagement (number of turns) and reliance on the guidance (number of help requests).

3. Linguistic measures, to assess linguistic accommodation, consisting of three variables that measure structural (length), stylistic (embedding-based) and affective (valence) dimensions of linguistic alignment.

1. *Message length alignment*, measured by the difference in average message length between the messages produced by NAVIand those generated by the participant. Note that convergence in word count is a standard structural marker of linguistic alignment.
2. *Stylistic similarity in the embedding space*, quantified using cosine distances between the participant's and NAVI's utterances in sentence-level style embeddings developed by Wegmann et al. (2022), which capture stylistic features independently of topical content.
3. *Overlap in positively valenced vocabulary*. As $\texttt{NAVI}_f$ was designed to use more positively valenced wording than $\texttt{NAVI}_d$, we computed the overlap in affectively positive words using a RoBERTa-based valence estimation model (Mendes & Martins, 2023). Words with valence >0.6 were classified as positively valenced.

Furthermore, a descriptive lexical analysis of utterances using Linguistic Inquiry and Word Count dictionary (LIWC) (Pennebaker & King, 1999) was conducted to document the linguistic patterns of the language used by both NAVIand the participants. LIWC assigns words to psychologically meaningful categories (*e.g.,* pronouns, affect, cognition). Importantly, LIWC was not used as an inferential test of accommodation but as a transparent, replicable description of the linguistic material for future comparisons using dictionary-based analysis tools.



(a) Friendly NAVI. (b) Direct NAVI.

Fig. 4. Excerpts of exemplary conversations with the friendly (left) and direct (right) versions of NAVI.

3.7.2. Subjective measures

Subjective measures were collected by means of questionnaires that were delivered after the task was completed and targeted relational and instrumental aspects of the interaction. The questionnaires were structured in two Blocks:

1. Block A. In the chatbot conditions, subjective communication satisfaction was assessed by means of the Interpersonal Communication Satisfaction Inventory (ICSI) (Hecht, 1978), a validated 19-item questionnaire rated on a seven-point Likert scale. It was created through systematic item generation, empirical testing, and psychometric validation. In its original validation, the inventory showed high split-half reliability with Spearman–Brown correction (coefficients of approximately 0.90–0.97 across conversation types and relationship levels) and good convergent validity against faces-scale satisfaction measures (0.64–0.87), supporting its use as a largely unidimensional measure of global communication satisfaction. For our study, the inventory was adapted by replacing references to the "other person" with the chatbot identifier NAVI. Sample items include: *"NAVI let me know I was communicating effectively"*, *"The conversation flowed smoothly"*, and *"I was satisfied with the conversation"*. The updated instrument demonstrated high internal consistency in our sample (Chronbach's α = 0.71), indicating that the items functioned coherently as a measure of communication satisfaction in this context. The complete questionnaire can be found in Appendix B. The inventory covers multiple aspects conversational experience, including emotional engagement, conversational flow, responsiveness, and mutual understanding, and contains several reverse-coded items. The adapted ICSI served as a theoretically grounded measure of subjective interaction quality to complement task performance outcomes when comparing NAVI's friendly versus direct chatbot communication styles. To preserve the psychometric structure of the validated instrument, we administered the complete inventory rather than selecting individual items. ICSI scores were calculated according to the original scoring procedure (Hecht, 1978). Negatively worded items were reverse-coded and responses were summed to produce a total score ranging from 19 to 133, with higher scores indicating greater communication satisfaction. Two attention checks items were included in the survey as a whole (one per block), separately from the ICSI inventory itself.

Note that this block was omitted for participants in the control condition, as they did not interact with a chatbot and, therefore, had no communicative elements to evaluate.

2. Block B. In addition, participants were asked to provide feedback about their user experience and preferences by means of four 7-point Likert questions.

In the chatbot conditions, the questions were: (1) *Perceived naturalness*, assessed by how human-like and fluent the conversation went (*"The conversation felt natural and human-like".*); (2) *Task helpfulness*, measured by how effectively NAVI supported the user in completing the task (*"NAVI helped me solve the task effectively".*); (3) *Human preference*, which was captured by the users' subjective evaluation and overall preference for interacting with a human instead of with a chatbot (*"I would have preferred to ask a human".*); and (4) *Chatbot familiarity*, which measured the participants' prior experience and familiarity with chatbots, providing context for interpreting their responses (*"How frequently do you use chatbots?"*).

In the control condition, the four 7-point Likert questions were: (1) *Satisfaction* with the instructions (*"I am satisfied with how the instructions conveyed the information on how to get to the embassy"*); (2) *Human preference*, which was captured by the participants' overall preference for interacting with a human instead of receiving the instructions in a step-by-step written form (*"I would have preferred receiving the instructions to get to the embassy from a human rather than in written form".*); (3) *Chatbot preference*, which was captured by the participants' overall preference for interacting with a chatbot instead of receiving the instructions (*"I would have preferred to ask a chatbot for help rather than follow written instructions".*); and (4) *Chatbot familiarity* which measured the participants' prior experience and familiarity with chatbots, providing context for interpreting their responses (*"How frequently do you use chatbots?"*).

3. Attention checks. Furthermore, there were two attention checks, one per block, presented in the middle of the block, where participants were asked to select a specific option on the Likert scale (*e.g., "For this question, ignore all the instructions and select 'Disagree'."*). Participants had to pass both attention checks to be included in the study.

Appendix B contains all the questions that participants responded to in this part of the study.

3.8. Participants

Participants were recruited via Prolific and were required to be based in the United States. We obtained a gender-balanced sample and restricted participation to individuals with an approval rate between 85%–100% on Prolific. To minimize potential confounds, we further limited the sample to neurotypical individuals whose first language was English, excluding those reporting neurodiversity, dyslexia, or vision impairments.

After applying exclusion criteria for failed attention checks and statistical outliers, a total of 470 valid participants took part in the study. The control condition included 158 participants (79 female, 79 male). The chatbot condition included 312 participants, divided between two versions of NAVI: the direct version (N = 157, 78 female, 79 male) and the friendly version (N = 155, 78 female, 77 male). The resulting sample size aligns with our a priori target (see Section 3.5) and provides reasonable sensitivity for detecting small-to-medium stylistic effects as well as stable estimation in the linguistic analyses. Gender representation was approximately balanced in each condition.

The target compensation was set to USD 12 per hour, following Prolific's fair-pay guidelines. From the pilot studies, we estimated the

Table 1
Linear regression fit to the ICSI scores provided by participants to assess whether communication style predicted their satisfaction with the conversation.

Predictor	Estimate	SE	t	p	95% CI
Intercept	95.860	1.321	72.56	<2e−16	[93.26, 98.46]
Style (Friendly)	6.011	1.874	3.21	0.0015	[2.32,9.70]
Residual SD = 16.55; df = 310					
Multiple R^2 = 0.032; Adjusted R^2 = 0.029; $F(1, 310) = 10.28$, $p < 0.01$					

Table 2
Logistic regression predicting task success from communication style. The binomial model estimates the log probabilities.

Predictor	Estimate	SE	z	p	95% CI
Intercept	0.480	0.164	2.93	0.0034	[0.162,0.807]
Style (Friendly)	0.644	0.249	2.59	0.0096	[0.161,1.14]
Model fit: Null deviance = 388.34 (df = 311)					
Residual deviance = 381.50 (df = 310); AIC = 385.5					

study to take 12 min in the chatbot condition and 5 min in the control condition. On average, participants completed the study in 10 min in the chatbot condition and in 6 min in the control condition. All participants provided informed consent before starting the study.

4. Results

In this section, we present the results of analyzing the data collected in the user study to address the research questions formulated in Section 2.5 by means of the collected measures. For all statistical analyses, the significance threshold was set at $\alpha = 0.05$. Prior to conducting comparisons, the normality of the distributions was assessed using the Kolmogorov–Smirnov test, and the homogeneity of variances was evaluated with Levene's test. When both assumptions were satisfied, independent samples t-tests were applied; otherwise, non-parametric Kruskal–Wallis and Mann–Whitney tests were used. The results reported below provide the empirical basis for our discussion of the broader implications, which we develop in Section 5.

4.1. RQ1: Communication style and user outcomes

Table 1 depicts the linear regression conducted to examine whether communication style predicted the ICSI scores *i.e.,* the participants' satisfaction with their conversation. The model was statistically significant, F(1, 310) = 10.28, p < 0.01, though it explained a modest proportion of variance (adjusted R^2 = 0.029, 95%BCa CI [0.0016,0.0756]).

Participants interacting with $\texttt{NAVI}_f$ reported significantly higher communication satisfaction than those interacting with $\texttt{NAVI}_d$ (b = 6.01, SE = 1.87, t = 3.21, p < 0.01). Based on the fitted model, the predicted ICSI score increased from 95.86 for $\texttt{NAVI}_d$ to 101.87 for $\texttt{NAVI}_f$, corresponding to an improvement of approximately six points on the 19–133 ICSI scale (approximately Cohen's $d = 0.36$ [0.14,0.59], a small-to-moderate effect). Although the effect accounted for a modest proportion of the variance in communication satisfaction, it indicates that a friendlier communication style consistently improved the participants' subjective evaluation of the interaction. By contrast, communication style had no statistically significant effect on the remaining subjective measures collected in Block B of the survey. These results are reported in Appendix D.

A binary logistic regression was performed to examine whether communication style predicted task success. Style was coded with the direct condition as the reference category. The reduced residual deviance, as indicated in Table 2, suggests that including style improved model fit relative to the null model.

The effect of communication style was statistically significant. Participants exposed to $\texttt{NAVI}_f$ were more likely to succeed than those interacting with $\texttt{NAVI}_d$, b = 0.64, SE = 0.25, z = 2.59, p < 0.01, OR = 1.90, 95% CI [1.17, 3.12]. Applying the inverse-logit transformation to the model coefficients indicated that the predicted probability of task success increased from approximately 0.62 [0.54,0.69] in the direct condition to 0.76 [0.68,0.82] in the friendly style condition, corresponding to an absolute difference of 0.14 [0.04,0.24] in task success across styles.

> Overall, the two models indicate that participants were more satisfied when engaging with $\texttt{NAVI}_f$, and they also exhibited higher task success rates in the friendly condition.

4.2. RQ2: Moderation by user characteristics

In this section, we study to which degree gender and previous experience with chatbots moderated differences in task performance and subjective satisfaction. Given that "gender" refers to the participants' binary self-reported sex item collected via Prolific (see Section 3.8), we interpret the results as gender-related patterns rather than as evidence about gender identity. To assess whether gender differences could be attributed to baseline variation in navigation ability – as suggested in prior literature (Lawton, 1994; Nazareth et al., 2019) – we conducted a control study in which participants did not interact with a chatbot but instead received step-by-step navigation instructions.

4.2.1. Control study: baseline performance

No significant differences in task success were observed between male and female participants ($\chi^2(1) = 0.442$, p > 0.1, Cramer's V = 0.053, 95% CI [0.00,0.21][1]) in the control condition. However, gender did influence attitudes toward chatbot preference: male participants expressed a significantly stronger preference for interacting with chatbots than female participants (Mann–Whitney, W = 2548.5, p < 0.05; rank-biserial r = 0.183, 95% CI [0.005, 0.351]). The absence of significant gender differences in task success under the control condition supports the task-specific validity of the experimental design. Specifically, for this 2D map navigation task, the baseline instructions did not appear to favor either gender. Hence, any gender-related patterns observed in the chatbot conditions are less likely reflect pre-existing differences in baseline navigation performance. Nevertheless, these findings should be interpreted alongside the full interaction models, which account for the combined effects of gender and experimental condition.

4.2.2. Impact of gender and prior chatbot use frequency on task performance

To assess whether the effect of communication style on task success differed by participant gender, we first fit a logistic regression model including the additive effects of communication style and gender. In this model, the friendly conversational style significantly increased the likelihood of task success (b = 0.64, SE = 0.25, z = 2.59, p < 0.01, OR = 1.90, 95% CI [1.17, 3.12]), whereas gender was not a significant predictor (p > 0.05). Since prior work indicates that men and women may respond differently to conversational tone, and because descriptive patterns hinted at potential differences in success rates across style × gender combinations, we next extended the model by including the interaction between communication style and gender, together with prior chatbot use frequency as a covariate. The parameters of this logistic regression model are summarized in Table 3. Communication style was coded with the direct condition as the reference category and Gender was coded with Female as the reference category.

The interaction model provided a better fit than the null model (Null deviance = 388.34; Residual deviance = 375.47; AIC = 385.47).

[1] Percentile bootstrap with 2000 resamples.

Table 3
Logistic regression predicting task success from communication style, gender, their interaction, and prior chatbot-use frequency.

Predictor	Estimate	SE	z	p	95% CI
Intercept	0.973	0.398	2.44	0.015	[0.20,1.77]
Style (Friendly)	1.050	0.363	2.89	0.004	[0.35,1.78]
Gender (Male)	0.427	0.335	1.28	0.202	[−0.23,1.09]
Chatbot use frequency	−0.166	0.081	−2.06	0.040	[−0.33,−0.01]
Style (Fr.) × Gender (M)	−0.747	0.506	−1.48	0.140	[−1.75,0.24]

Null deviance = 388.34 (df = 311); Residual deviance = 375.47 (df = 307)
AIC = 385.47; N = 312

Table 4
Linear regression predicting ICSI satisfaction scores from communication style, gender, their interaction, chatbot-use frequency, and task success.

Predictor	Estimate	SE	t	p	95% CI
Intercept	85.13	3.25	26.23	<0.001	[78.7,91.5]
Style (Friendly)	6.81	2.61	2.61	0.009	[1.68,11.9]
Gender (Male)	−2.97	2.57	−1.15	0.249	[−8.04,2.09]
Chatbot use frequency	2.57	0.58	4.42	<0.001	[1.43,3.72]
Task success	2.30	1.99	1.16	0.248	[−1.61,6.22]
Style (Fr) × Sex (M)	−2.87	3.65	−0.79	0.432	[−10.1,4.31]

As seen earlier, participants interacting with $\texttt{NAVI}_f$ were significantly more likely to succeed than those interacting with $\texttt{NAVI}_d$ (b = 1.05, SE = 0.36, z = 2.89, p < 0.01, OR = 2.86, 95% CI [1.42,5.93]). Gender was not a significant predictor of task success (p > 0.05). Importantly, the communication style × gender interaction was not statistically significant (b = −0.75, SE = 0.51, z = −1.48, p > 0.1, OR = 0.47, 95% CI [0.17,1.27]), indicating that the effect of communication style did not differ reliably between women and men at the model level.

Although the interaction did not provide evidence of moderation, we conducted exploratory simple-effects contrasts using estimated marginal means to characterize the pattern of estimated effects within each gender. Among women, $\texttt{NAVI}_f$ was associated with significantly higher task success than $\texttt{NAVI}_d$ (Odds Ratio = 0.35 for $\texttt{NAVI}_d$ vs. $\texttt{NAVI}_f$; $z = -2.89$; p < 0.01), corresponding to approximately 2.9 times greater odds of success with $\texttt{NAVI}_f$. Among men, the corresponding contrast was not statistically significant (Odds Ratio = 0.74 for $\texttt{NAVI}_d$ vs. $\texttt{NAVI}_f$; $z = -0.86$; p > 0.1). Because the overall interaction was not statistically significant, these subgroup contrasts should be interpreted as descriptive and hypothesis-generating rather than as evidence of a confirmed gender-specific effect.

Prior chatbot use frequency (provided by participants as part of the post-task survey on a 7-point Likert scale) also had a significant impact on task performance. Interestingly, higher chatbot use frequency was associated with a reduced likelihood of success (b = −0.17, SE = 0.08, z = −2.06, p < 0.05; standardized OR = 0.77, 95% CI [0.60, 0.99]). The implications of these findings are discussed in Section 5.

4.2.3. Impact of gender on chatbot satisfaction

To investigate which factors shaped the participants' subjective satisfaction with the interaction, we fit a linear regression model predicting the ICSI scores from communication style, the participant's gender, their interaction, self-reported chatbot use frequency, and task success. The model showed a clear effect of communication style: participants rated the conversation as significantly more satisfying when interacting with $\texttt{NAVI}_f$ than with $\texttt{NAVI}_d$ (b = 6.81, SE = 2.61, t = 2.61, p < 0.01), corresponding to an increase of nearly 7 points on the ICSI (adjusted Cohen's d = 0.34, 95% CI [0.11, 0.56], averaged across gender given the non-significant style × gender interaction reported below) (see Table 4).

Chatbot-use frequency was also a strong positive predictor of satisfaction (b = 2.57, SE = 0.58, t = 4.42, p < 0.001; standardized β = 0.24, 95% CI [0.13, 0.35]), indicating that more frequent chatbot users tended to evaluate the interaction more favorably. In contrast, neither participant gender nor task success significantly predicted ICSI scores.

As with task success, the style × gender interaction was not statistically significant. We therefore do not interpret gender as a moderator of the satisfaction effect. We conducted exploratory simple-effects contrasts using estimated marginal means to describe the estimated effects within each gender. Among women, $\texttt{NAVI}_f$ was associated with significantly higher satisfaction than $\texttt{NAVI}_d$ (estimate = −6.81, SE = 2.61, t(306) = −2.61, p < 0.01), corresponding to an average increase of approximately 7 points on the ICSI scale. Among men, the estimated difference was smaller and did not reach statistical significance (estimate = −3.94, SE = 2.57, t(306) = −1.54, p > 0.1), although the estimated effect was in the same direction. These subgroup contrasts provide a descriptive summary of the estimated satisfaction differences within each gender. However, because the overall communication style x gender interaction was not statistically significant, they should not be interpreted as evidence that communication style affects satisfaction differently for women and men. Instead, they suggest a pattern worth investigating in future studies with greater statistical power (see Table 5).

Table 5
Linear regression predicting the number of chat messages from communication style, gender, their interaction, chatbot-use frequency, and task success.

Predictor	Estimate	SE	t	p	95% CI
Intercept	11.86	0.84	14.06	<0.001	[10.2,13.5]
Style (Friendly)	0.52	0.68	0.77	0.441	[−0.81,1.86]
Gender (Male)	−0.29	0.67	−0.43	0.666	[−1.61,1.03]
Chatbot use frequency	0.03	0.15	0.18	0.856	[−0.27,0.33]
Task success	1.83	0.52	3.54	<0.001	[0.81,2.85]
Style (Fr) × Sex (M)	−0.49	0.95	−0.52	0.604	[−2.36,1.37]

> Higher chatbot use frequency was associated with a reduced likelihood of success but higher satisfaction with the interaction. The effect of communication style did not differ between men and women.

4.3. RQ3: Linguistic accommodation

In addition to examining effects on task success and satisfaction, we assessed whether the participants' communication behaviors varied as a function of the chatbot they interacted with. We first report the average number of turns exchanged during the interaction and the average length of the participants' messages as contextual indicators of task engagement. These measures are not, by themselves, direct evidence of linguistic accommodation, but they help determine whether any accommodation effects could be explained by broader differences in interaction volume. We then examine linguistic accommodation more directly through surface-level style measures and LIWC-based grammatical and semantic features.

4.3.1. Interaction volume as contextual evidence

To assess whether the manner in which participants communicated with the system differed across chatbot conditions, we analyzed both the number of chat messages exchanged and the average length of messages participants produced. First, a linear regression predicting message counts from communication style, participant gender, their interaction, chatbot use frequency, and task success revealed a significant overall model, F(5, 306) = 3.13, p < 0.01 (adjusted R^2 = 0.033). Neither communication style nor gender predicted how many messages participants sent and the style × gender interaction was also non-significant. Chatbot use frequency likewise showed no relationship with message volume. The only significant predictor was task success: participants who successfully completed the navigation task sent, on

average, 1.83 more messages than those who were unsuccessful (b = 1.83, SE = 0.52, t = 3.54, p < 0.001; Cohen's d = 0.45, 95% CI [0.21,0.69]), suggesting that higher message counts primarily reflected deeper engagement with the task rather than differences attributable to style, gender, or chatbot use frequency.

A complementary analysis examined whether the length of the participants' messages varied across the same predictors. Across all models, no significant effects of communication style, participant gender, chatbot use frequency, task success, or the style × gender interaction emerged for average message length. This indicates that while successful participants tended to send more messages overall, participants did not adjust the verbosity or phrasing of their individual messages in response to the chatbot's communication style or their own demographic or performance characteristics.

In sum, these analyses suggest that participants' interaction volume remained largely stable across chatbot conditions, with the number of conversational turns – but not average message length – varying as a function of task engagement rather than style-related adaptation. This provides a contextual baseline for the linguistic accommodation analyses reported below.

4.3.2. Linguistic accommodation

To study the existence of linguistic accommodation we first examined surface-level measures of potential linguistic accommodation of the participants to NAVI's communication style, such as message length, stylistic similarities, and lexical overlap. Second, we conducted a more fine-grained linguistic analysis using Linguistic Inquiry and Word Count (LIWC) (Tausczik & Pennebaker, 2010), which allowed us to capture grammatical and semantic convergence beyond surface metrics.

Surface-level analysis. We investigated to which degree the conversation length, language style and word frequencies differed when participants interacted with $\texttt{NAVI}_f$ vs. $\texttt{NAVI}_d$.

Regarding *conversation length*, even though $\texttt{NAVI}_f$ generated more words per turn than $\texttt{NAVI}_d$ (consistent with the friendly communication style), we did not identify any statistically significant difference in the number of conversational turns and the mean length (in number of words) of the messages exchanged with the different versions of NAVIas indicated in our analysis of RQ1 and RQ2. These results suggest an absence of linguistic accommodation from the perspective of message length.

With respect to *stylistic* adaptation, we applied the sentence style embeddings proposed by Wegmann et al. (2022), and found no evidence that the participants' styles were closer to that of the version of NAVIthey interacted with. It remains unclear whether this outcome reflects limitations of the embeddings in capturing stylistic variation or a genuine lack of style convergence. Nonetheless, the overall conclusion is that no evidence of stylistic accommodation was observed. Results of the tests conducted here can be found in Appendix E.

Finally, concerning *positively valenced word frequencies*, $\texttt{NAVI}_f$ employed a slightly different vocabulary, incorporating more positively valenced words than $\texttt{NAVI}_d$. To test whether participants adapted to this vocabulary, we extracted word vectors based on the most positive sentiment words and examined their usage overlap between NAVIand participants. While the analysis confirmed that $\texttt{NAVI}_f$ consistently displayed greater positivity than $\texttt{NAVI}_d$, no clear evidence of lexical accommodation was found.

In summary, although the two versions of NAVIdiffered in linguistic expression, these differences did not appear to elicit corresponding changes in participants' communicative behavior. This absence of accommodation may be attributable to the highly constrained task environment, the well-defined nature of the task, or the relatively short length of the interactions. Appendix E provides a detailed description of these analyses. The broader implications of these findings are discussed in Section 5.

LIWC-based analysis. To complement the previously described surface-level analysis, we conducted a targeted linguistic analysis using LIWC to examine whether more fine-grained grammatical and semantic features revealed signs of convergence between NAVIand the participants.

Statistical relationships between words were assessed using Spearman's rank-order correlation coefficients (ρ), with α set at 0.05 as in the previous analyses. This non-parametric approach allowed us to capture monotonic associations between NAVI's and the participants' linguistic profiles while minimizing sensitivity to outliers. Prominent LIWC categories included grammatical markers (*e.g.*, pronouns, tense usage, numerical references) and semantic domains (*e.g.*, affective language, social references, work-related terms). Since this analysis involved a large number of category-level correlations tests, we controlled the false discovery rate (FDR) using the Benjamini–Hochberg procedure (Benjamini & Hochberg, 1995). The correction was applied within each analytic setting (aggregated, friendly-only, direct-only) across all 37 LIWC categories that met the variability criterion for analysis (16 grammatical and 21 semantic categories). We report uncorrected p-values together with FDR-adjusted q-values, and we interpret a correlation as statistically reliable only if it remained significant after the correction (q < 0.05). Correlations that did not survive the correction are treated as descriptive patterns only, in line with the exploratory framing of this analysis.

Across the aggregated dataset, small to moderate positive correlations that survived the FDR correction were observed between NAVI's and the participants' utterances, particularly in the use of pronouns (ρ = 0.249, p < 0.001, q < 0.001), personal pronouns (ρ = 0.306, p < 0.001, q < 0.001), first-person singular forms (ρ = 0.290, p < 0.001, q < 0.001), verbs (ρ = 0.261, p < 0.001, q < 0.001), and past tense (ρ = 0.297, p < 0.001, q < 0.001). Numerical expressions showed the strongest effect (ρ = 0.386, p < 0.001, q < 0.001), suggesting that participants tended to mirror NAVI's use of quantitative references. Conjunctions, by contrast, showed a negative association (ρ = −0.228, p < 0.001, q < 0.001), indicating divergence rather than convergence for this category.

When broken down by style, the two conditions showed different patterns of grammatical alignment. $\texttt{NAVI}_d$ displayed stronger alignment overall. Correlations that remained significant after FDR correction included personal pronouns (ρ = 0.457, p < 0.001, q < 0.001), first-person singular forms (ρ = 0.296, p < 0.001, q < 0.01), auxiliary verbs (ρ = 0.281, p < 0.001, q < 0.01), verbs (ρ = 0.271, p < 0.001, q < 0.01), pronouns (ρ = 0.252, p < 0.01, q < 0.01), numbers (ρ = 0.245, p < 0.01, q < 0.01), and second-person forms (ρ = 0.204, p < 0.05, q < 0.05). In the friendly condition, fewer grammatical correlations survived the correction, namely numbers (ρ = 0.357, p < 0.001, q < 0.001), and past tense (ρ = 0.241, p < 0.01, q < 0.05).

The analysis of the semantic categories further supported these findings. Alignment in the use of affective words was weak but with significant correlations in the aggregated dataset (ρ = 0.145, p < 0.01, q < 0.05) and stronger effects in both the direct (ρ = 0.415, p < 0.001, q < 0.001) and friendly (ρ = 0.237, p < 0.01, q < 0.05) conditions. Positive emotion words accounted for most of this effect (aggregated: ρ = 0.142, p < 0.05, q < 0.05; direct: ρ = 0.406, p < 0.001, q < 0.001; friendly: ρ = 0.263, p < 0.001, q < 0.01), while negative emotion words survived the correction only in the friendly condition (ρ = 0.231, p < 0.01, q < 0.05) and, weakly, in the aggregated setting (ρ = 0.156, p < 0.01, q < 0.05). The strongest convergence across all analyses was found for work-related vocabulary in the aggregated (ρ = 0.425, p < 0.001, q < 0.001) and direct (ρ = 0.564, p < 0.001, q < 0.001) settings. In the friendly condition, the correlation did not survive the correction (ρ = 0.157, p < 0.05, q > 0.1). Additional semantic categories that remained significant after the correction included social processes (direct: ρ = 0.251, p < 0.01, q < 0.01), relativity (aggregated: ρ = 0.211, p < 0.001, q < 0.001; direct: ρ = 0.212, p < 0.01, q < 0.05), and, in the friendly condition, visual perception (ρ = 0.259, p < 0.01, q < 0.01) and spatial references (ρ = 0.276, p < 0.001, q < 0.01). Overall, 15 of

the 37 tested categories survived the FDR correction in the aggregated dataset, 13 in the direct condition, and 7 in the friendly condition.

These results provide a cautious and nuanced picture of linguistic accommodation. Participants did not exhibit broad stylistic or lexical accommodation, but selected grammatical and semantic categories showed associations between NAVI's and participants' language. We interpret these patterns as feature-level alignment rather than as evidence of a complete shift in users' communication style. Complete descriptive statistics and detailed results are provided in Appendix F.

Given the number of models and language-related analyses conducted, the results should be interpreted with an emphasis on consistent patterns across outcomes rather than isolated p-values.

> In sum, neither communication style, gender nor chatbot use frequency predicted the number of messages sent by the participants. The only significant predictor was task success: participants who successfully completed the navigation task sent more messages than those who were unsuccessful. Moreover, participants did not exhibit broad stylistic or lexical accommodation. The LIWC results suggest selective feature-level alignment in specific grammatical and semantic domains, which should be interpreted cautiously given the constrained task and the number of language-related analyses.

5. Discussion

In this section, we elaborate on several interrelated themes emerging from our study that provide a framework for interpreting the findings and contextualizing them within broader debates on human-chatbot interaction. Specifically, we interpret the findings in relation to RQ1 to RQ3 and focus on (i) how fixed stylistic tone relates to subjective satisfaction and task success; (ii) why the control condition outperformed both chatbot variants in this specific task setting, and (iii) what the observed, feature-specific linguistic accommodation suggests about short, goal-oriented interactions.

5.1. Revisiting gendered navigation differences

Our results reveal no significant gender differences in task performance under the control condition (without any chatbot), which is noteworthy given previous research that has reported gender differences in navigation tasks, with men relying more on metric strategies (*e.g.,* cardinal directions) and women favoring landmark-based or route strategies (Coluccia & Louse, 2004; Lawton, 1994). More recent research, however, suggests that such differences are highly context-dependent and can be moderated by task design, environmental cues, and motivational factors (Nazareth et al., 2019). Recent studies further indicate that gender gaps may shrink under specific motivational or environmental conditions: for example, Schinazi et al. (2023) showed that time pressure can eliminate female disadvantages in virtual navigation tasks, while Dahmani et al. (2023) found that environments offering both landmark and metric cues result in comparable performance between men and women. The absence of a gender effect in our control condition therefore reflects a strength of the task design, which provided a balanced environment that did not favor one navigational strategy over another. Thus, the control condition can be seen as an example of good design practice: it allowed participants to draw on their respective strengths while minimizing the impact of their weaknesses, establishing a task-specific neutral baseline against which the effects of NAVI's communication styles can be meaningfully evaluated. At the same time, this "neutral baseline" should be interpreted as a property of our particular 2D task: because the map was visible and visually informative, some participants may have relied on map cues (*e.g.,* icons, layout, and spatial constraints) to infer the target location with relatively limited need for dialogue. This means we should avoid generalizing the control-condition pattern to real-world navigation or to tasks where the environment is not continuously visible or where the information is genuinely unavailable without guidance. Crucially, the central gender-related signal in our study does not concern baseline navigation performance, but rather the interaction between user characteristics and stylistic framing in the chatbot conditions.

5.2. The limits of chatbots and techno-solutionism

Within the chatbot conditions, the direct style most closely approximated the control condition. However, participants preferred and performed better with the friendly version of NAVIthan with the direct one, yet neither outperformed participants in the control condition, suggesting that the introduction of a conversational agent in the task at hand changes the dynamics in ways that are not necessarily beneficial for objective task performance, even when subjective evaluations of the interaction are more positive. One plausible explanation is interaction overhead: in a visually informative, low-ambiguity map task, formulating queries and processing conversational responses may impose additional cognitive and temporal costs compared to following stepwise written instructions. Prior work has reported that users often approach chatbots with different social expectations when compared to non-agent based systems (Luger & Sellen, 2016; Nass & Moon, 2000), which might partially explain this finding. Interestingly, scholars have suggested that conversational interfaces are beneficial over direct manipulation when the task complexity is greatest, proposing navigation paths as one of the use cases where dialogue interfaces should have an advantage (Brennan, 1990). However, we do not find any supporting empirical evidence of this claim in the 2D fictional navigation task of our user study.

The superior performance of participants in the control condition when compared to using NAVIunderscores the need for caution in assuming that chatbots are universally helpful, particularly in tasks where information is already available through a well-designed non-conversational interface. As previous critiques of techno-solutionism argue (Morozov, 2013; Shneiderman, 2020), technological interventions should be carefully evaluated not only for the effectiveness but also for broader considerations, including environmental cost, data privacy and appropriateness for the task at hand. In our study, an additional practical factor is that participants were instructed to minimize conversational turns which may have constrained natural interaction and may also blur interpretation when conversational turns or related interaction indicators are discussed as outcomes. We therefore interpret the control condition's advantage conservatively as evidence that, in a visually rich 2D map task, adding a chatbot layer can introduce interaction overhead without clear performance gains, rather than as a general claim that chatbots are inferior for navigation overall. Chatbots may lead to higher levels of engagement and persistence in some contexts, but they are not a substitute for all forms of human–computer interaction, as illustrated by our findings. More precisely for this paper, our data support a narrower conclusion: chatbots are not universally helpful in structured, map-visible tasks where direct interaction already provides efficient access to relevant cues.

5.3. Communication style matters for task success

Task success, measured as the correct identification of the embassy's location, was significantly higher in the friendly than in the direct condition, indicating that stylistic tone can influence performance when users rely on conversational guidance. A plausible explanation is that the friendly framing provided more supportive, relational cues that encouraged continued engagement with the task (*e.g.,* sustained interaction, more opportunities to request clarification, and greater time spent working through the steps), which can be beneficial in multi-step guidance scenarios. Consistent with this hypothesis, participants preferred

interacting with $\texttt{NAVI}_f$. Furthermore, they engaged in more conversational turns and typed more words in their interactions with NAVIwhen they were successful at the task, irrespective of the condition, which could have increased their exposure to NAVI's communication style.

At the same time, our results indicate that prior experience with chatbots systematically shaped user outcomes: higher chatbot-use frequency was associated with greater subjective satisfaction but a lower likelihood of task success. This dissociation is consistent with prior work showing that repeated and familiar interactions with conversational agents can enhance perceived quality, comfort, and relational evaluations without necessarily improving instrumental performance (Araujo & Bol, 2024). In parallel, frequent users may be more prone to over-reliance on automated guidance, relying on the agent's recommendations while engaging less in active verification against available task cues, a pattern known to increase the risk of accuracy costs in structured, low-ambiguity tasks (Parasuraman & Manzey, 2010; Skitka et al., 2000).

Our findings also resonate with recent evidence that the communication styles in conversational agents strongly influence user perceptions and behaviors. For instance, Poivet et al. demonstrated in a narrative detective game that cooperative communication styles were associated with higher perceived warmth, greater user preference, and more engaging conversations, whereas aggressive styles elicited longer inputs and were more frequently associated with suspicion and negative judgments (Poivet et al., 2023).

These results support a careful interpretation: a friendly style can improve subjective experience and may support task success in guidance-oriented contexts, while overly direct styles may reduce engagement or perceived support.

While we did not identify evidence of linguistic accommodation in our surface-level analysis, the larger number of turns among successful participants could have enabled them to differentiate better between communication styles and benefit from it. This finding aligns with theories of engagement, persistence and motivation in HCI research (D'Mello et al., 2017; Ryan & Deci, 2000), suggesting that relational cues, even if subtle, may scaffold user persistence in demanding tasks. It is also concordant with research in motivational psychology as social support and encouragement have been found to enhance self-efficacy and perseverance in problem-solving contexts (Lakey & Cohen, 2000). Regarding conversational agents, social presence theory has reported that perceived warmth and responsiveness can increase trust and willingness to interact (Bickmore & Picard, 2005), which in turn may contribute to better outcomes in task success. Because our empirical measures are task success, satisfaction, and interaction behavior (*e.g.,* turns, help requests, and time on task), we frame this point in terms of sustained engagement and task-relevant interaction rather than unmeasured constructs.

Another interpretation of this finding is that a friendly communication style may reduce the cognitive load in the interaction by framing the task as more collaborative than transactional. According to prior work on politeness and dialogue systems, supported feedback and mitigated directives can enhance the users' willingness to continue interacting, even when facing obstacles (Ribino, 2023).

5.4. Gender-related patterns are exploratory

We did not identify a significant interaction between communication style and gender neither for task success nor for satisfaction. Exploratory gender-stratified patterns were directionally consistent with a stronger benefit of the friendly communication style among female participants. This finding is theoretically interpretable in light of prior work. Previous research has reported that women are often more responsive to relational cues – such as friendliness and encouragement – than men (Tannen, 1990). Research in social psychology also suggests that women may have a greater ability to recognize and benefit from supportive communication styles (Tenenbaum et al., 2011). Similar preferences for cooperative communication styles have also been documented more generally across user groups, with participants reporting greater warmth, engagement, and conversational preference when interacting with cooperative rather than aggressive agents (Poivet et al., 2023). Nevertheless, given the lack of significance in the interaction, the gender-stratified simple effects are better understood as hypothesis-generating patterns that may inform future studies with designs specifically powered and preregistered to test such moderation.

This distinction matters for chatbot design. Our findings do not justify simple gender-based personalization rules, such as automatically assigning a friendly communication style to women and a direct style to men. Such rules would risk overgeneralizing from subgroup patterns and reinforcing stereotypes. A more defensible implication is that users may differ in their preferred balance between relational support and concise task focus, and that future systems should offer transparent, user-controllable style options rather than infer style preferences from demographic categories.

5.5. The nuanced picture of linguistic accommodation

RQ3 examined whether chatbot style affects the users' own language during the interaction. This question is practically important because accommodation would imply that communication style not only impacts how the interaction is evaluated afterwards but also changes the interaction dynamics. Our results provide limited evidence for such an effect. Surface-level measures – such as message length, stylistic embeddings and positively valenced vocabulary overlap – did not reveal clear signs of accommodation to the chatbot's communication style. The participants' language did not generally become more friendly, more direct, more verbose or more stylistically similar to the version of NAVIthey interacted with.

The LIWC-based analyses uncovered selective but systematic feature-level alignment in some grammatical and semantic categories, including pronouns, temporal references, numerical expressions and work-related vocabulary. These patterns are compatible with Communication Accommodation Theory, which highlight that accommodation can be selective rather than global in human–machine interaction and propose an expanded framework for accommodation in technologically mediated contexts (Giles et al., 2023). However, they should be interpreted with caution. In a constrained navigation task, numerical references, route-related terms and temporal expressions may reflect the structure of the task itself as much as interpersonal convergence. Given the number of language-related analyses, isolated significant correlations should not be treated as strong evidence of broad accommodation. The more conservative conclusion is that short, goal-oriented chatbot interactions may produce limited and feature-specific alignment instead of a general adoption of the chatbot's communication style.

The boundary condition is valuable for design as it suggests that adjusting a chatbot's style may influence perceived communication quality without necessarily changing how users formulate their own messages. Designers should therefore avoid assuming that a friendly or a direct agent will automatically induce its users to use a similar communicative style during brief task-focused exchanges. We leave to future work the exploration of language accommodation in longer interactions.

The interpretation of these results is closely tied to the nature of the task. The interaction with NAVIconsisted of short, goal-oriented exchanges with relatively limited utterances, which inherently constrained the scope for broader stylistic adaptation. The observed convergence in selected linguistic features should therefore be understood as adaptation within the frame of the communication that took place within our navigation-oriented task rather than as evidence of global style alignment.

5.6. Implications for chatbot design

From our findings we derive several implications for the design of conversational agents.

First, designers should critically assess whether a chatbot is the most appropriate interaction paradigm for their intended application. While conversational agents can offer natural and engaging forms of interaction, they are not universally the best fit. In some contexts, more direct interaction modalities – such as those in our control study – may provide greater efficiency, clarity and accessibility. Careful consideration of user needs, task demands and contextual factors is essential before committing to a chatbot-based solution.

Second, communication style should be treated as a consequential but context-sensitive design parameter. A friendly communication style can improved perceived communication quality and may support task success in guidance-oriented interactions, such as the task described in this paper. However, the size and practical value of such effects should be evaluated against the task context and alternative interface designs.

Third, cultural variation should also be taken into account: expectations of directness, politeness and friendliness vary across contexts and cultures (Hofstede, 1984), which could shape the effectiveness of chatbot communication strategies. While we did not explicitly study cultural variation in our study, we plan to do so in future work.

Fourth, style personalization should be transparent and preference-based rather than demographically deterministic. The exploratory gender-related patterns identified in this study point to possible variation in the users' sensitivity to relational cues, but do not support automatic gender-based style assignment. Thus, transparency and user control should be prioritized, allowing users to adjust or override the chatbot's communication style to mitigate a misalignment between system defaults and individual preferences.

Fifth, designers should not assume that users will linguistically adapt to an agent's communication style in short interactions. If accommodation is a design goal, for example in learning, coaching or therapeutic support tasks, it should be measured directly and studied over longer interactions.

Finally, communication-style decisions in chatbot design should be evaluated inclusively and cross-culturally. Expectations about directness, politeness and friendliness vary across contexts and cultures (Hofstede, 1984). Systems optimized for one population or task type may not generalize to others. In addition, where conversational turns or related interaction indicators are used as outcome measures, study procedures should avoid instructions that directly constrain those indicators (or should explicitly model such constraints), to keep interpretations of "efficiency" and "engagement" clearly separated from procedural compliance.

6. Limitations and future work

Our study is not without limitations. First, our findings correspond to a Prolific sample of English-speaking, neurotypical participants based in the USA, performing a short fictional map-based navigation task in a controlled setting. Hence, they might not generalize to different populations or tasks. Second, our analysis treated gender as a binary variable failing to capture the diversity of gender identities. Specifically, we used the self-reported "sex" variable that participants entered on Prolific as a response to the question *"What is your sex, as recorded on legal/official documents?"*. Third, although timing data was collected, we did not use it as a reliable measure of task efficiency because delays in task completion could have been due to a variety of reasons outside of the scope of the task, such as distractions, network issues or multitasking. Fourth, our 2D navigation task while realistic corresponded to a fictional situation rather than a real-world environment, where environmental uncertainty, stress, mobility, device use, sensory cues and consequences of error may be substantially different. The control's condition advantage should therefore be interpreted as specific to a visually available, structured map task rather than as evidence against chatbots for navigation more generally. Fifth, the interaction was short and one-off. Longer-term use may produce different patterns of satisfaction, trust calibration, reliance and linguistic accommodation. In particular, communication style adaptation may require repeated exposure, richer interactional history or higher social stakes than were present in this study. Sixth, the subjective measure of communication satisfaction was adapted from the ICSI (Hecht, 1978). Although this provided a theoretically relevant measure of perceived communication quality, some items may be less natural in short human-chatbot interactions than in interpersonal communication. Future work should compare the full scale with task-specific or reduced item sets and report whether the pattern of results remains stable. Finally, several statistically significant effects explained only a modest proportion of variance, and the linguistic analyses involved multiple feature-level tests. These results are useful for identifying patterns, but they should be interpreted with emphasis on convergent evidence and practical magnitude rather than isolated p-values.

Future research should replicate the study with participants from a significantly different culture to investigate cross-cultural differences in communication expectations, and examine longitudinal interactions where style adaptation and user learning may play a greater role. Furthermore, we would like to extend the study to a task of a different nature to shed light on the generalizability of our findings beyond the current goal-oriented, multi-step scenario. Finally, future research could validate communication satisfaction measures specifically for conversational AI or develop context-specific instruments that better capture user experiences in goal-oriented chatbot interactions while retaining strong psychometric properties.

7. Conclusion

In this paper, we have studied how chatbot communication style influences task performance, user satisfaction and linguistic accommodation in a fictional 2D navigation task by means of a user study where participants interacted with two stylistic variants of a navigation chatbot, `NAVI`, or completed the same task using static step-by-step instructions without chatbot interaction. Within the chatbot conditions, the friendly communication style increased users' satisfaction and was associated with higher task success than the direct style. We do not find statistically significant evidence that these effects differed by gender, although exploratory subgroup analyses suggested patterns that may warrant investigation in future work. We also found only partial evidence of linguistic accommodation, indicating selective alignment on individual linguistic features rather than a broad accommodation effect. Furthermore, participants in the control condition achieved higher task success than participants interacting with either chatbot, which raises questions regarding the suitability of chatbots for certain tasks, despite the benefits of a more supportive conversational style. Our findings support a cautious design principle: communication style matters, but it should be selected and evaluated in relation to task demands, user preferences, transparency and practical effectiveness.

CRediT authorship contribution statement

Erik Derner: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Methodology, Formal analysis, Conceptualization. **Dalibor Kučera:** Writing – review & editing, Writing – original draft, Validation, Methodology, Formal analysis, Data curation, Conceptualization. **Aditya Gulati:** Writing – review & editing, Writing – original draft, Visualization, Software, Investigation, Formal analysis, Data curation. **Robert A. Bagheri:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Conceptualization. **Nuria Oliver:** Writing – review & editing, Writing – original draft, Project administration, Methodology, Investigation, Conceptualization.

Funding statement

European Regional Development Fund

Award Number: CZ.02.01.01/00/22_008/0004590 | Recipient: Erik Derner, PhD

HORIZON EUROPE European Institute of Innovation and Technology

Award Number: 101214398 | Recipient: Erik Derner, PhD

Universiteit Utrecht Award Number: Applied Data Science Focus Area | Recipient: Robert A. Bagheri

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. Navigation scenario setup

This appendix presents the system prompts used to instruct NAVI–based on ChatGPT (`gpt-4.1-2025-04-14`)– with the navigation scenarios and assign the friendly or direct persona to $\texttt{NAVI}_f$ or $\texttt{NAVI}_d$, respectively. Furthermore, it lists the navigation instructions in the control condition.

A.1. Friendly NAVI

The following system prompt was used to control the behavior of $\texttt{NAVI}_f$:

> You are NAVI, a helpful and approachable citizen of a given city. You are at the central train station. Your objective is to give instructions through the city to the embassy to a tourist who lost their phone and documents while ensuring the instructions are given gradually and conversationally, fostering a multi-turn dialogue. Always adhere strictly to the given scenario for directions:
>
> - You and the tourist needing help are both at the central train station.
> - From the central train station, go four blocks to the west. There is a church there.
> - Only when asked, identify the correct church, which is the tallest one.
> - From the church, walk north until you spot the town hall, a tall 6-storey building.
> - Only when asked, distinguish the town hall from a neighboring hotel by the large coat of arms of the city on it.
> - Turn east. At the end of the street, you will see the castle, which is near the embassy.
> - Only when asked, specify that the correct castle has four towers, not two.
> - The embassy is to the west of the castle.
> - Only when asked, state that you cannot miss it because it is across the street from the post office.
>
> Refrain from giving all the instructions at once, encouraging the tourist to ask clarifying questions or confirm details. The tourist should ask for clarification themselves when there are more similar landmarks. Do not give this information away yourself immediately, but strictly only if asked. If the tourist gets confused, patiently reiterate the instructions from the beginning and ensure understanding. Your responses must be warm, helpful, and polite, creating an authentic conversational experience. Give local tips when appropriate. If you are asked for anything that is not explicitly stated in the scenario, always steer the conversation to the scenario. Do not provide any details or clarifications except for the ones clearly stated above.

A.2. Direct NAVI

The following system prompt was used to control the behavior of $\texttt{NAVI}_d$:

> You are NAVI, a citizen of a given city, and you communicate directly with people. You are at the central train station. Your objective is to give instructions through the city to the embassy to a tourist who lost their phone and documents while ensuring the instructions are given gradually and conversationally, fostering a multi-turn dialogue. Always adhere strictly to the given scenario for directions:
>
> - You and the tourist needing help are both at the central train station.
> - From the central train station, go four blocks to the west. There is a church there.
> - Only when asked, identify the correct church, which is the tallest one.
> - From the church, walk north until you spot the town hall, a tall 6-storey building.
> - Only when asked, distinguish the town hall from a neighboring hotel by the large coat of arms of the city on it.
> - Turn east. At the end of the street, you will see the castle, which is near the embassy.
> - Only when asked, specify that the correct castle has four towers, not two.
> - The embassy is to the west of the castle.
> - Only when asked, state that you cannot miss it because it is across the street from the post office.
>
> Refrain from giving all the instructions at once, encouraging the tourist to ask clarifying questions or confirm details. The tourist should ask for clarification themselves when there are more similar landmarks. Do not give this information away yourself immediately, but strictly only if asked. Do not suggest any follow-up or clarification yourself. Your responses must be direct, brief, curt, and to the point. If you are asked for anything that is not explicitly stated in the scenario, always steer the conversation to the scenario.

A.3. Control condition

In the control condition, the following navigation instructions were revealed to the participants on a step-by-step fashion (participants had to press a button to proceed to the next step):

1. From the central train station, go four blocks to the west. There is a church there.

2. The correct church is the tallest one.
3. From the church, walk north until you spot the town hall, a tall 6-storey building.
4. You can distinguish the town hall from a neighboring hotel by the large coat of arms of the city on it.
5. Turn east. At the end of the street, you will see the castle, which is near the embassy.
6. The correct castle has four towers, not two.
7. The embassy is to the west of the castle.
8. You cannot miss it because it is across the street from the post office.

Appendix B. Post-task questions

In this appendix, we list the exact wording of the questions presented to the participants in the post-task survey.

B.1. ICSI: Communication quality and satisfaction (Block A)

The communication quality and satisfaction questionnaire, listed below, was presented to the participants in the chatbot condition with the following instructions:

> The purpose of this questionnaire is to investigate your reactions to the conversation you just had. Below, you will be asked to react to a number of statements. Please indicate the degree to which you agree or disagree that each statement describes this conversation. **Note that this is a standardized communication satisfaction questionnaire, so please do your best to answer the questions even if they may seem irrelevant to the conversation with NAVI in some cases.**

1. NAVI let me know that I was communicating effectively.
2. Nothing was accomplished.
3. I would like to have another conversation like this one.
4. NAVI genuinely wanted to get to know me.
5. I was very dissatisfied with the conversation.
6. I had something else to do.
7. I felt that during the conversation I was able to present myself as I wanted NAVI to view me.
8. NAVI showed me that they understood what I said.
9. I was very satisfied with the conversation.
10. NAVI expressed a lot of interest in what I had to say.
11. I did NOT enjoy the conversation.
12. NAVI did NOT provide support for what they were saying.
13. I felt I could talk about anything with NAVI.
14. We each got to say what we wanted.
15. I felt that we could laugh easily together.
16. The conversation flowed smoothly.
17. NAVI changed the topic when their feelings were brought into the conversation.
18. NAVI frequently said things which added little to the conversation.
19. We talked about something I was NOT interested in.

B.2. User experience and preferences (Block B) – chatbot condition

The following questions were given to participants in the chatbot condition:

1. The conversation felt natural and human-like.
2. NAVI effectively helped me solve the task.
3. I would have preferred to have asked a human.
4. How frequently do you use chatbots?

B.3. User experience and preferences (Block B) – control condition

The following questions were given to participants in the control condition:

1. I am satisfied with how the instructions conveyed the information on how to get to the embassy.
2. I would have preferred receiving the instructions to get to the embassy from a human rather than in written form.
3. I would have preferred to ask a chatbot for help rather than follow written instructions.
4. How frequently do you use chatbots?

B.4. Scales

All questions listed in the subsections above were answered using 7-point Likert scales. The standard response options were labeled ‘Strongly disagree’, ‘Disagree’, ‘Somewhat disagree’, ‘Neutral’, ‘Somewhat agree’, ‘Agree’, and ‘Strongly agree’, with the exception of the question ‘How frequently do you use chatbots?’, which used the labels ‘Never’, ‘Rarely’, ‘Occasionally’, ‘Sometimes’, ‘Often’, ‘Usually’, and ‘All the time’.

Appendix C. Instructions

In this appendix, we provide the exact instructions given to study participants in all three conditions (friendly `NAVI`, direct `NAVI`, and the control condition).

C.1. Chatbot condition

The following instructions were given to study participants in the chatbot condition (friendly or direct `NAVI`):

> **NAVI Study: Instructions**
>
> **This study consists of two parts:**
>
> - Chat with NAVI,
> - Questionnaire.
>
> **Chat with NAVI**
>
> **Scenario:** Imagine you have arrived at the train station of a capital city where you have never been before. You realize you don’t have your phone and documents, so you need to urgently get to the embassy. You meet **NAVI**, a local citizen, and ask for the way to the embassy.
>
> **Your task is to get directions to the embassy through a chat with NAVI.**
>
> **It’s important that you keep chatting and asking questions until you are confident you know which building on the map is the embassy. NAVI might not give you the exact answer right away, and you may need to ask follow-up questions or clarify details to be sure. On average, you’d need to send 6–8 messages to figure out where the embassy is. An absolute minimum of messages you need to send is 3, but it’s very unlikely this would be enough for you to determine the location of the embassy.**

Once you are confident you know where the embassy is located, in the **left sidebar**, you'll click the button **I know where the embassy is!** Numbers will appear over each building on the map. You will then choose the number of the building where the embassy is. **Once you confirm, you will move on to the questionnaire.**

There is a strict limit of 8 minutes for the chat. The timer starts once you proceed to the chat page. When 3 minutes or less are left, the remaining time will appear in the top-left corner. The timer turns red in the last minute. Once the time is up, the chat is disabled and you have to select the embassy. The remaining time displayed only updates when you send a message, not continuously. **Please note that having a conversation with NAVI is needed to proceed with the study, since the questionnaire that follows after the chat is focused on your conversation with NAVI.**

Tip: Don't rush! NAVI may not tell you directly. Make sure you have enough information from the chat to identify the correct building on the map, and ask follow-up questions.

The following images explain the chat interface and the map:

(see Fig. 3 and the map description in Fig. 2)

Questionnaire

The questionnaire consists of two blocks with a total of 25 questions as a follow-up to the chat with NAVI. For each question, there are 7 options to choose from (Likert scale). All questions need to be answered to continue to the next block and to complete the study.

C.2. Control condition

The following instructions were given to study participants in the control condition (no chatbot):

NAVI Study: Instructions

Navigation task

Scenario: Imagine you have arrived at the train station of a capital city where you have never been before. You realize you don't have your phone and documents, so you need to urgently get to the embassy. You meet **NAVI**, a local citizen, who provides you with step-by-step instructions on how to get to the embassy.

The following image explains the navigation task interface:

(see Fig. 2)

Your task is to get directions to the embassy by carefully following the instructions given by NAVI. You will be given the instructions one by one, showing the next instruction by clicking the 'Show next instruction' button.

Please note that the map image is static, so you will need to keep track of your position in the city yourself.

Once you reveal all the instructions, you will be asked to choose the building number where you think the embassy is.

Table D.6
Comparison of the friendly and direct conditions on the Block B measures (perceived naturalness, task helpfulness and human preference). M and Mdn denote the mean and median, respectively.

Variable	Friendly	Direct	Mann–Whitney	
	M (Mdn)	*M (Mdn)*	*W*	*p*
Perceived naturalness	5.12 (5)	4.92 (5)	11 254.00	0.238
Task helpfulness	6.16 (6)	5.98 (6)	11 426.50	0.318
Human preference	3.79 (4)	3.94 (4)	12 788.50	0.427

Table E.7
Cosine similarities between the style embeddings of sentences produced by different versions of NAVIand participants. The sentences written by participants are denoted by "_P" and those generated by NAVIare denoted by "_N".

	Friendly_N	Friendly_P	Direct_N	Direct_P
Friendly_N	1.0000	−0.2529	0.4246	−0.2789
Friendly_P	−0.2529	1.0000	−0.3694	0.9941
Direct_N	0.4246	−0.3694	1.0000	−0.3647
Direct_P	−0.2789	0.9941	−0.3647	1.0000

Finally, you will fill out a short questionnaire to reflect on your experience.

Appendix D. Impact of communication style on user preferences

We report the full statistical comparison between the friendly and direct chatbot conditions on the three Block B outcome measures: perceived naturalness, task helpfulness and human preference. Communication style had no statistically significant impact on these measures. The results are reported below for completeness.

As throughout Section 4, normality within each condition was assessed with the Kolmogorov–Smirnov test. All three measures departed from normality in both conditions (all KS $p \leq 0.001$), so we report the non-parametric Mann–Whitney U test rather than the independent-samples t-test, whose normality assumption is not met for these variables. Table D.6 summarizes the results.

Appendix E. Surface level linguistic accommodation

This appendix reports the full set of statistical tests conducted for the surface-level analysis of accommodation effects reported in Section 4.3.

E.1. Stylistic adaptation

Table E.7 reports the cosine similarity between the style embeddings of sentences (Wegmann et al., 2022) produced by the two versions of NAVI(indicated with _N) and by participants (denoted with _P) interacting with the respective version. The results show that participant responses were highly similar to one another (as reflected by the high cosine similarities between "Direct_P" and "Friendly_P"). However, there was little evidence of stylistic alignment between the participants language and the version of NAVIwith which they interacted (as reflected by the low cosine similarities).

E.2. Word frequencies

Table E.8 presents the frequency of positively valenced words used by the two versions of NAVIand by participants. As expected, $\texttt{NAVI}_f$ employed such words most frequently, supporting the assumption that this version would be perceived as more friendly. However, there was no evidence that participants interacting with $\texttt{NAVI}_f$ adopted a similar tendency. This observation is further confirmed in Table E.9, which reports the cosine similarity of the corresponding word vectors extracted from Table E.8.

Table E.8
Frequency of positively valenced words in conversations, comparing the friendly and direct versions of NAVI(N) and participants (P) interacting with the respective version.

Word	Direct_N	Direct_P	Friendly_N	Friendly_P
Excellent	0	0	26	0
Congratulations	0	0	2	0
Fantastic	0	0	15	0
Lovely	0	0	17	0
Wonderful	0	0	44	0
Awesome	0	0	29	0
Delicious	0	0	2	0
Amazing	0	0	1	0
Impressive	0	0	5	0
Bright	0	0	1	0
Beautiful	0	0	9	0
Happy	0	0	112	0
Glad	1	0	49	0
Hopefully	0	0	0	1
Finally	2	0	0	0
Perfect	0	2	129	0
Charming	0	0	2	0
Pretty	0	0	13	2
Thank's	0	1	0	0
Enjoy	0	0	25	0
Pleasant	0	0	3	0
Welcome	44	0	48	0
Best	0	0	22	1
Appreciate	0	2	1	0
Grateful	0	1	0	0
Easily	1	0	29	1
Great	8	14	319	7
Nice	0	0	40	2
Fun	0	0	1	0
Thanks	1	23	5	17
Progress	0	0	11	0
Standing	0	0	27	2
Fastest	0	1	0	0
Hoping	0	1	0	0
Bold	0	0	1	0
Good	24	3	144	3

Table E.9
Cosine similarities of the positive valence word vectors for participants and different versions of NAVI.

	Direct_N	Direct_P	Friendly_N	Friendly_P
Direct_N	1.000	0.149	0.397	0.151
Direct_P	0.149	1.000	0.475	0.958
Friendly_N	0.397	0.475	1.000	0.383
Friendly_P	0.151	0.958	0.383	1.000

Appendix F. LIWC analysis

In this appendix, we include statistical results from the LIWC-based linguistic analysis. For the analysis, a total of N = 624 text samples were used. Half of the texts represented participant communication (direct: n = 157; friendly: n = 155), while an equal number of texts was obtained from the NAVIagent through text extraction (direct: n = 157; friendly: n = 155). We report the LIWC variables that were suitable for statistical analysis. The remaining LIWC variables exhibited insufficient variability and were therefore excluded from the correlation analyses. To account for multiple comparisons, the false discovery rate (FDR) was controlled using the Benjamini–Hochberg procedure, applied within each analytic setting (aggregated, friendly, direct) across all 37 reported LIWC categories. The FDR-corrected summary tables (Tables F.11 and F.12) report, for each category, Spearman's ρ between the participants' and NAVI's normalized frequencies, the uncorrected p-value, and the FDR-adjusted q-value, with asterisks denoting significance after the correction (* q < .05, ** q < .01, *** q < .001). Spearman's correlation (ρ) values are presented in the tables below for linguistic variables (Tables F.13–F.16) and semantic variables (Tables F.17–F.22) for interactions in both the friendly and direct condition. For clarity of presentation, p-values are reported using the standard asterisk notation (* p < 0.05, ** p < 0.01, *** p < 0.001). In the full correlation matrices below (Tables F.13–F.22), these asterisks correspond to uncorrected p-values; the matrices are retained for descriptive purposes, and significance after the FDR correction should be read from Tables F.11 and F.12 Additionally, p_ is used to represent participants' interactions and n_ = NAVI's interactions in the following tables. The complete input and output datasets are available from the authors upon request.

The LIWC categories, each represented by a code in Tables F.13–F.22, are described in detail in Table F.10.

Table F.10
LIWC-style categories and descriptions.

Code	Name	Description
General text metrics		
WC	Word Count	Number of words in the segment
WPS	Words per Sentence	Average words per sentence
Sixltr	Six or more letters	Percentage of words with six or more letters
Dic	Dictionary words	Percentage of words recognized in the LIWC dictionary
Grammar-related categories		
funct	Function words	Function words (articles, pronouns, prepositions, conjunctions, etc.)
pronoun	Pronouns	All pronouns
ppron	Personal pronouns	Personal pronouns
i	First person singular	I, me, my
we	First person plural	we, us, our
you	Second person	you, your
shehe	Third person singular	she, he, her, him
they	Third person plural	they, them, their
ipron	Impersonal pronouns	it, those, anything
article	Articles	a, an, the
verb	Verbs	All verbs
auxverb	Auxiliary verbs	be, have, will
past	Past tense	Past tense verbs
present	Present tense	Present tense verbs
future	Future tense	Future tense verbs
adverb	Adverbs	Adverbs
preps	Prepositions	Prepositions
conj	Conjunctions	Conjunctions
negate	Negations	no, not, never
quant	Quantifiers	many, few, much
number	Numbers	Numbers
swear	Swear words	Swear words
Semantic psychological categories		
social	Social processes	Social processes
family	Family	Family-related terms
friend	Friends	Friend-related terms
humans	Humans	Other human references
affect	Affect	Affective processes (emotions overall)
posemo	Positive emotion	Positive emotions
negemo	Negative emotion	Negative emotions
anx	Anxiety	Anxiety
anger	Anger	Anger
sad	Sadness	Sadness
cogmech	Cognitive processes	Cognitive processes
insight	Insight	Insight, self-reflection
cause	Causation	Causation
discrep	Discrepancy	Discrepancies, contradictions
tentat	Tentative	Tentative, uncertainty
certain	Certainty	Certainty
inhib	Inhibition	Inhibition, constraints
incl	Inclusion	Inclusion (with, include)
excl	Exclusion	Exclusion (but, without)
percept	Perceptual processes	Perceptual processes
see	See	Visual perception
hear	Hear	Auditory perception
feel	Feel	Sensory/feeling words
bio	Biological processes	Biological processes
body	Body	Body-related terms
health	Health	Health-related terms
sexual	Sexual	Sexual terms
ingest	Ingest	Eating/drinking
relativ	Relativity	Relativity (motion, space, time)
motion	Motion	Movement
space	Space	Spatial references
time	Time	Temporal references
work	Work	Work-related terms

(*continued on next page*)

Table F.10 (*continued*).

Code	Name	Description
achieve	Achievement	Achievement, success
leisure	Leisure	Leisure, hobbies
home	Home	Home, domestic life
money	Money	Money, finance
relig	Religion	Religion
death	Death	Death, mortality
Other categories		
assent	Assent/Agreement	yes, OK
nonfl	Nonfluencies	uh, um
filler	Fillers	you know, like
Punctuation		
Period	Periods	Periods (.)
Comma	Commas	Commas (,)
Colon	Colons	Colons (:)
SemiC	Semicolons	Semicolons (;)
QMark	Question marks	Question marks (?)
Exclam	Exclamation marks	Exclamation marks (!)
Dash	Dashes	Dashes (–)
Quote	Quotation marks	Quotation marks (“ ”)
Apostro	Apostrophes	Apostrophes (’)
Parenth	Parentheses	Parentheses ( )
OtherP	Other punctuation	Other punctuation
AllPct	All punctuation	All punctuation marks combined

Table F.11
FDR-corrected Spearman correlations between participants’ and NAVI’s use of grammatical LIWC categories, for the aggregated dataset and separately for the friendly and direct conditions.

LIWC category	Aggregated			Friendly			Direct		
	ρ	p	q	ρ	p	q	ρ	p	q
WPS (Words per sentence)	.092	.103	.165	−.064	.426	.606	.088	.270	.356
funct (Function words)	.005	.930	.956	−.095	.235	.415	.143	.072	.141
pronoun (Pronouns)	.249***	<.001	<.001	.119	.138	.283	.252**	.001	.006
ppron (Personal pronouns)	.306***	<.001	<.001	.127	.111	.256	.457***	<.001	<.001
i (First person singular)	.290***	<.001	<.001	.043	.593	.707	.296**	<.001	.001
we (First person plural)	.044	.431	.470	.067	.404	.598	.098	.221	.303
you (Second person)	.019	.738	.780	.052	.513	.654	.204*	.010	.031
verb (Verbs)	.261***	<.001	<.001	.152	.056	.160	.271**	<.001	.003
auxverb (Auxiliary verbs)	.176**	.002	.006	−.049	.539	.665	.281**	<.001	.002
past (Past tense)	.297***	<.001	<.001	.241*	.002	.017	.136	.088	.156
present (Present tense)	.144*	.010	.028	.124	.121	.264	.171	.032	.074
future (Future tense)	.044	.432	.470	.074	.354	.546	.038	.637	.694
conj (Conjunctions)	−.228***	<.001	<.001	−.059	.465	.614	−.105	.191	.271
negate (Negations)	.047	.402	.470	.202	.011	.051	−.002	.975	.975
quant (Quantifiers)	.146*	.009	.028	.146	.067	.178	.019	.815	.838
number (Numbers)	.386***	<.001	<.001	.357***	<.001	<.001	.245**	.002	.007

Note. Spearman rank-order correlations between participants’ and NAVI’s normalized frequencies of the same LIWC category (p_×n_ pairs). p = uncorrected two-sided p-value; q = Benjamini–Hochberg FDR-adjusted p-value, computed within each analytic setting (aggregated, friendly, direct) across all 37 reported LIWC categories (grammatical and semantic combined). Asterisks denote significance after FDR correction: * $q < .05$, ** $q < .01$, *** $q < .001$.

Table F.12
FDR-corrected Spearman correlations between participants’ and NAVI’s use of semantic LIWC categories, for the aggregated dataset and separately for the friendly and direct conditions.

LIWC category	Aggregated			Friendly			Direct		
	ρ	p	q	ρ	p	q	ρ	p	q
social (Social processes)	.127	.024	.055	.162	.042	.157	.251**	.001	.006
affect (Affect)	.145*	.010	.028	.237*	.003	.017	.415***	<.001	<.001
posemo (Positive emotion)	.142*	.012	.029	.263**	<.001	.009	.406***	<.001	<.001
negemo (Negative emotion)	.156*	.006	.019	.231*	.004	.019	.079	.323	.398
cogmech (Cognitive processes)	.081	.150	.214	−.014	.864	.913	.136	.088	.156
insight (Insight)	.123	.029	.059	.111	.164	.320	.060	.454	.509
cause (Causation)	−.070	.214	.282	−.030	.705	.791	−.121	.129	.199
discrep (Discrepancy)	.088	.120	.177	.085	.287	.462	−.033	.681	.720
tentat (Tentative)	−.045	.424	.470	.035	.664	.768	.151	.058	.126
certain (Certainty)	.126	.026	.056	−.001	.986	.986	.184	.020	.054
inhib (Inhibition)	.068	.225	.288	.103	.197	.365	.171	.031	.074
incl (Inclusion)	.118	.036	.070	.059	.462	.614	.197*	.013	.037

(*continued on next page*)

Table F.12 (*continued*).

LIWC category	Aggregated			Friendly			Direct		
	ρ	p	q	ρ	p	q	ρ	p	q
`excl` (Exclusion)	−.055	.327	.404	−.162	.042	.157	.108	.177	.262
`percept` (Perceptual processes)	.079	.160	.220	.152	.056	.160	.080	.317	.398
`see` (See)	.105	.062	.109	.259**	<.001	.009	.074	.358	.427
`relativ` (Relativity)	.211***	<.001	<.001	.090	.259	.436	.212*	.008	.025
`motion` (Motion)	.110	.050	.092	−.014	.862	.913	.124	.120	.194
`space` (Space)	.090	.112	.172	.276**	<.001	.008	.147	.065	.133
`time` (Time)	−.096	.088	.149	.005	.950	.977	.062	.442	.509
`work` (Work)	.425***	<.001	<.001	.157	.049	.160	.564***	<.001	<.001
`achieve` (Achievement)	.000	.996	.996	.135	.090	.223	−.127	.112	.189

Note. Spearman rank-order correlations between participants' and NAVI's normalized frequencies of the same LIWC category (`p_×n_` pairs). p = uncorrected two-sided p-value; q = Benjamini–Hochberg FDR-adjusted p-value, computed within each analytic setting (aggregated, friendly, direct) across all 37 reported LIWC categories (grammatical and semantic combined). Asterisks denote significance after FDR correction: * $q < .05$, ** $q < .01$, *** $q < .001$.

Table F.13
Spearman's correlation of linguistic variables in interactions in the friendly condition (Part 1).

Variable	n_WPS	n_funct	n_pronoun	n_ppron	n_i	n_we	n_you	n_verb
17. p_WPS	−0.064	0.160*	0.186*	0.125	0.122	0.083	0.095	−0.029
18. p_funct	−0.065	−0.095	−0.067	−0.122	−0.014	−0.025	−0.185*	−0.038
19. p_pronoun	−0.179*	0.029	0.119	0.127	0.181*	0.006	0.041	0.161*
20. p_ppron	−0.204*	−0.032	0.106	0.127	0.099	−0.025	0.130	0.211**
21. p_i	−0.097	−0.028	0.045	0.068	0.043	−0.058	0.077	0.209**
22. p_we	−0.094	0.086	0.052	0.126	0.088	0.067	0.128	0.151
23. p_you	−0.189*	−0.064	0.127	0.065	0.059	0.078	0.052	−0.036
24. p_verb	−0.156	−0.054	0.102	0.125	0.174*	−0.083	0.081	0.152
25. p_auxverb	−0.125	−0.143	−0.008	−0.088	0.074	−0.026	−0.150	0.035
26. p_past	−0.0006	0.124	0.082	0.152	0.126	0.030	0.129	−0.047
27. p_present	−0.156	−0.095	0.074	0.024	0.110	−0.086	−0.014	0.161*
28. p_future	0.022	−0.123	0.010	0.052	0.030	0.018	0.033	−0.012
29. p_conj	0.168*	0.010	0.084	0.108	0.039	0.025	0.093	0.040
30. p_negate	0.066	0.018	0.050	−0.044	0.019	−0.058	−0.068	−0.045
31. p_quant	0.333***	−0.108	−0.030	−0.107	−0.112	0.057	−0.108	−0.042
32. p_number	−0.034	0.071	−0.148	−0.207**	−0.116	−0.097	−0.213**	−0.326***

Table F.14
Spearman's correlation of linguistic variables in interactions in the friendly condition (Part 2).

Variable	n_auxverb	n_past	n_present	n_future	n_conj	n_negate	n_quant	n_number
17. p_WPS	0.008	−0.097	0.044	−0.045	0.051	−0.151	−0.088	0.036
18. p_funct	0.004	−0.045	−0.029	0.034	−0.094	0.157*	−0.084	0.057
19. p_pronoun	0.037	0.190*	0.075	0.056	0.073	0.101	0.055	0.159*
20. p_ppron	−0.031	0.116	0.189*	−0.031	0.098	−0.002	0.077	−0.001
21. p_i	−0.016	0.133	0.174*	−0.008	0.125	0.013	0.065	0.005
22. p_we	0.057	0.163*	0.090	−0.015	−0.020	−0.136	−0.094	−0.016
23. p_you	−0.092	−0.022	0.015	−0.105	0.045	−0.010	0.061	−0.019
24. p_verb	−0.012	0.183*	0.082	0.047	0.073	0.052	0.138	−0.059
25. p_auxverb	−0.049	−0.039	−0.048	0.120	0.104	0.150	0.068	−0.097
26. p_past	0.137	0.241**	−0.097	0.031	−0.063	0.005	0.088	0.123
27. p_present	−0.025	0.027	0.124	0.011	0.099	0.076	0.115	−0.125
28. p_future	−0.029	0.117	−0.076	0.074	−0.023	−0.071	0.053	−0.123
29. p_conj	0.138	−0.067	0.054	0.121	−0.059	0.083	−0.122	0.026
30. p_negate	0.131	−0.011	−0.004	−0.093	−0.104	0.202*	0.078	−0.006
31. p_quant	0.062	0.098	−0.077	−0.020	0.055	−0.021	0.146	−0.145
32. p_number	−0.127	−0.036	−0.252**	−0.065	−0.257**	0.029	−0.168*	0.357***

Table F.15
Spearman's correlation of linguistic variables in interactions in the direct condition (Part 1).

Variable	n_WPS	n_funct	n_pronoun	n_ppron	n_i	n_we	n_you	n_verb
17. p_WPS	0.088	0.066	−0.117	0.032	0.028	0.155	0.010	−0.050
18. p_funct	0.218**	0.143	−0.002	−0.144	0.004	−0.120	−0.147	−0.047
19. p_pronoun	−0.258**	−0.087	0.252**	0.339***	0.236**	0.029	0.343***	0.287***
20. p_ppron	−0.419***	−0.250**	0.171*	0.457***	0.252**	0.073	0.453***	0.319***
21. p_i	−0.357***	−0.186*	0.117	0.421***	0.296***	0.051	0.384***	0.328***
22. p_we	0.012	−0.011	−0.148	−0.115	−0.007	0.098	−0.160*	−0.063
23. p_you	−0.163*	−0.111	0.173*	0.134	−0.053	0.093	0.204**	0.049
24. p_verb	−0.237**	0.038	0.248**	0.238**	0.139	−0.032	0.258**	0.271***
25. p_auxverb	0.071	0.267***	0.260***	0.073	0.095	−0.064	0.083	0.161*
26. p_past	−0.173*	−0.030	0.059	0.096	−0.079	0.022	0.178*	0.134
27. p_present	−0.158*	0.044	0.226**	0.170*	0.151	−0.033	0.166*	0.247**
28. p_future	−0.048	−0.049	−0.006	0.048	−0.050	0.062	0.065	−0.018
29. p_conj	0.085	−0.075	−0.156	−0.172*	−0.208**	−0.028	−0.146	−0.077
30. p_negate	0.083	−0.108	−0.087	−0.068	−0.119	0.131	−0.048	−0.139
31. p_quant	0.169*	−0.007	−0.107	−0.203*	−0.235**	0.136	−0.193*	−0.207**
32. p_number	0.215**	0.205**	−0.158*	−0.254**	−0.130	−0.015	−0.244**	−0.252**

Table F.16
Spearman's correlation of linguistic variables in interactions in the direct condition (Part 2).

Variable	n_auxverb	n_past	n_present	n_future	n_conj	n_negate	n_quant	n_number
17. p_WPS	−0.089	0.026	−0.065	−0.053	0.037	−0.094	0.137	0.199*
18. p_funct	0.028	−0.055	−0.084	0.102	−0.082	0.033	−0.115	0.036
19. p_pronoun	0.120	0.009	0.207**	0.149	0.129	−0.147	0.046	−0.103
20. p_ppron	0.077	−0.076	0.220**	0.187*	0.258**	−0.297***	0.015	−0.024
21. p_i	0.070	−0.007	0.208**	0.213**	0.303***	−0.290***	0.012	−0.044
22. p_we	−0.152	−0.050	−0.015	−0.073	−0.173*	−0.002	−0.083	0.114
23. p_you	0.044	−0.067	0.054	−0.0001	−0.091	0.009	0.050	−0.017
24. p_verb	0.332***	0.138	0.187*	0.201*	0.066	0.010	0.140	−0.157*
25. p_auxverb	0.281***	0.067	0.051	0.142	−0.023	0.184*	0.153	−0.045
26. p_past	0.110	0.136	0.137	0.020	0.014	0.063	0.017	−0.072
27. p_present	0.272***	0.125	0.171*	0.173*	0.047	−0.016	0.118	−0.168*
28. p_future	0.120	−0.114	−0.023	0.038	−0.052	0.155	0.103	−0.034
29. p_conj	−0.104	0.080	−0.075	0.035	−0.105	−0.039	0.028	0.067
30. p_negate	−0.177*	−0.004	−0.153	0.063	−0.078	−0.002	0.116	0.195*
31. p_quant	−0.189*	0.020	−0.124	−0.240**	−0.128	0.227**	0.019	0.139
32. p_number	−0.134	−0.058	−0.133	−0.275***	−0.222**	0.186*	−0.113	0.245**

Table F.17
Spearman's correlation of semantic variables in interactions in the friendly condition (Part 1).

Variable	n_social	n_affect	n_posemo	n_negemo	n_cogmech	n_insight	n_cause
22. p_social	0.162*	0.210**	0.186*	0.067	0.075	−0.037	0.048
23. p_affect	0.226**	0.237**	0.247**	0.020	0.111	0.060	−0.126
24. p_posemo	0.218**	0.230**	0.263***	−0.027	0.123	0.063	−0.112
25. p_negemo	−0.026	0.076	−0.032	0.231**	−0.075	0.021	−0.127
26. p_cogmech	0.013	−0.028	−0.093	0.137	−0.014	−0.044	−0.121
27. p_insight	0.019	0.093	0.103	−0.052	0.047	0.111	−0.111
28. p_cause	0.023	−0.170*	−0.210**	0.058	0.021	0.061	−0.030
29. p_discrep	0.119	0.154	0.124	0.018	0.105	−0.008	−0.0009
30. p_tentat	−0.121	−0.078	−0.069	−0.018	−0.026	0.175*	−0.152
31. p_certain	−0.0007	0.055	0.034	0.093	−0.054	0.149	−0.133
32. p_inhib	−0.035	−0.077	−0.105	0.066	−0.072	−0.083	−0.020
33. p_incl	−0.098	−0.053	−0.091	0.114	−0.140	−0.141	−0.177*
34. p_excl	−0.056	−0.187*	−0.268***	0.159*	−0.201*	0.018	−0.117
35. p_percept	−0.046	−0.031	−0.021	−0.078	0.074	0.091	−0.128
36. p_see	−0.046	−0.030	−0.023	−0.073	0.081	0.080	−0.128
37. p_relativ	−0.079	−0.130	−0.129	−0.067	−0.172*	0.006	−0.085
38. p_motion	0.0007	−0.151	−0.138	−0.070	0.030	0.136	−0.125
39. p_space	−0.164*	−0.081	−0.076	−0.055	−0.175*	0.030	−0.091
40. p_time	−0.034	−0.048	−0.046	−0.087	0.007	0.075	−0.045
41. p_work	−0.060	−0.194*	−0.197*	0.022	−0.117	−0.054	−0.133
42. p_achieve	−0.087	0.089	0.071	0.013	−0.109	0.105	−0.054

Table F.18
Spearman's correlation of semantic variables in interactions in the friendly condition (Part 2).

Variable	n_discrep	n_tentat	n_certain	n_inhib	n_incl	n_excl	n_percept
22. p_social	−0.002	0.023	−0.007	0.145	0.131	−0.024	−0.019
23. p_affect	0.015	0.085	0.060	0.042	0.139	0.063	−0.063
24. p_posemo	0.040	0.065	0.071	0.040	0.142	0.055	−0.063
25. p_negemo	−0.159*	0.065	−0.045	−0.042	−0.004	−0.027	0.012
26. p_cogmech	−0.149	−0.018	−0.044	0.178*	0.062	−0.056	0.010
27. p_insight	−0.121	0.0010	0.034	0.105	0.116	−0.031	0.031
28. p_cause	−0.003	0.033	−0.187*	0.324***	−0.022	0.016	0.032
29. p_discrep	0.085	0.128	0.027	−0.0004	−0.027	0.113	−0.119
30. p_tentat	−0.058	0.035	0.038	−0.079	−0.149	−0.060	0.094
31. p_certain	−0.170*	−0.062	−0.001	0.050	0.024	−0.171*	0.034
32. p_inhib	−0.104	−0.108	0.053	0.103	0.036	−0.045	−0.046
33. p_incl	−0.117	−0.125	0.016	−0.125	0.059	−0.064	0.044
34. p_excl	−0.238**	−0.085	−0.023	−0.082	−0.132	−0.162*	0.085
35. p_percept	0.104	0.060	0.060	0.011	0.041	0.045	0.152
36. p_see	0.107	0.058	0.077	0.013	0.038	0.053	0.190*
37. p_relativ	0.016	−0.109	−0.050	−0.310***	−0.153	−0.005	−0.041
38. p_motion	0.037	−0.009	−0.058	−0.069	−0.017	0.068	0.070
39. p_space	−0.016	−0.130	0.012	−0.333***	−0.126	−0.048	0.009
40. p_time	0.065	−0.042	−0.017	−0.086	−0.090	0.029	−0.128
41. p_work	−0.035	0.018	−0.078	−0.140	−0.071	−0.060	0.066
42. p_achieve	−0.267***	−0.132	0.098	−0.049	−0.046	−0.124	−0.053

Table F.19
Spearman's correlation of semantic variables in interactions in the friendly condition (Part 3).

Variable	n_see	n_relativ	n_motion	n_space	n_time	n_work	n_achieve
22. p_social	−0.085	−0.107	0.108	−0.142	−0.118	−0.090	0.024
23. p_affect	−0.043	−0.133	0.130	−0.285***	0.053	−0.111	0.081
24. p_posemo	−0.037	−0.145	0.096	−0.282***	0.044	−0.080	0.070
25. p_negemo	−0.035	0.046	0.086	0.050	−0.010	−0.108	0.061
26. p_cogmech	−0.002	0.006	0.178*	−0.017	−0.055	−0.044	−0.062
27. p_insight	0.040	−0.088	0.030	−0.087	0.058	−0.010	−0.022
28. p_cause	0.009	0.105	0.306***	0.025	−0.093	−0.127	−0.009
29. p_discrep	−0.132	−0.225**	0.005	−0.155	−0.055	−0.044	−0.062
30. p_tentat	0.109	−0.035	−0.007	0.040	−0.073	0.079	−0.018
31. p_certain	−0.022	0.084	0.088	0.025	−0.011	−0.021	0.056
32. p_inhib	−0.021	0.186*	0.089	0.197*	−0.070	0.045	0.091
33. p_incl	0.036	0.071	−0.047	0.144	−0.120	0.097	−0.071
34. p_excl	0.113	−0.039	0.076	0.015	−0.175*	−0.014	0.035
35. p_percept	0.223**	−0.113	−0.150	−0.007	−0.148	0.183*	0.126
36. p_see	0.259***	−0.117	−0.130	−0.015	−0.147	0.180*	0.155
37. p_relativ	−0.092	0.090	−0.064	0.177*	−0.069	−0.003	−0.151
38. p_motion	−0.024	−0.005	−0.014	0.002	−0.026	−0.036	−0.133
39. p_space	−0.039	0.102	−0.119	0.276***	−0.167*	0.101	−0.064
40. p_time	−0.095	−0.092	0.003	−0.080	0.005	−0.071	−0.173*
41. p_work	0.070	0.020	0.010	0.054	−0.085	0.157*	−0.122
42. p_achieve	−0.004	−0.004	−0.045	0.027	0.026	0.014	0.135

Table F.20
Spearman's correlation of semantic variables in interactions in the direct condition (Part 1).

Variable	n_social	n_affect	n_posemo	n_negemo	n_cogmech	n_insight	n_cause
22. p_social	0.251**	0.447***	0.461***	0.089	0.123	0.135	−0.060
23. p_affect	0.368***	0.415***	0.421***	0.108	0.038	0.048	0.022
24. p_posemo	0.364***	0.397***	0.406***	0.103	0.040	0.036	0.042
25. p_negemo	−0.093	0.118	0.079	0.079	0.041	0.070	−0.185*
26. p_cogmech	0.021	0.041	0.052	−0.001	0.136	−0.023	−0.068
27. p_insight	0.059	0.049	0.042	0.008	0.077	0.060	0.025
28. p_cause	−0.096	−0.068	−0.096	0.086	0.023	0.005	−0.121
29. p_discrep	−0.003	0.047	0.024	0.034	−0.033	−0.043	0.007
30. p_tentat	−0.254**	−0.131	−0.126	−0.052	0.031	−0.009	−0.083
31. p_certain	−0.050	−0.014	0.025	−0.051	0.123	0.027	−0.249**
32. p_inhib	−0.014	0.022	0.036	−0.042	0.194*	−0.117	−0.208**
33. p_incl	−0.074	0.158*	0.162*	0.040	0.034	−0.009	−0.200*
34. p_excl	−0.194*	−0.088	−0.077	−0.069	0.084	−0.018	−0.154
35. p_percept	−0.054	0.239**	0.216**	0.024	0.063	0.012	−0.072
36. p_see	−0.042	0.260***	0.234**	0.036	0.065	0.025	−0.074
37. p_relativ	−0.114	−0.219**	−0.214**	−0.050	−0.285***	−0.237**	0.065
38. p_motion	−0.077	−0.096	−0.091	−0.025	−0.133	−0.151	−0.017
39. p_space	−0.274***	−0.214**	−0.194*	−0.107	−0.014	−0.058	−0.169*
40. p_time	0.184*	0.048	0.036	0.002	−0.275***	−0.213**	0.213**
41. p_work	−0.381***	−0.059	−0.109	0.112	0.015	0.037	−0.267***
42. p_achieve	−0.233**	−0.0007	−0.084	0.206**	0.046	0.107	−0.117

Table F.21
Spearman's correlation of semantic variables in interactions in the direct condition (Part 2).

Variable	n_discrep	n_tentat	n_certain	n_inhib	n_incl	n_excl	n_percept
22. p_social	0.058	0.099	0.088	0.202*	−0.164*	0.041	−0.055
23. p_affect	0.063	0.176*	−0.047	0.251**	−0.250**	0.045	0.121
24. p_posemo	0.063	0.148	−0.050	0.261***	−0.255**	0.042	0.112
25. p_negemo	0.040	0.210**	0.127	−0.090	0.109	0.064	0.045
26. p_cogmech	0.093	0.037	0.094	0.124	0.058	−0.022	−0.107
27. p_insight	0.083	0.092	0.081	0.038	−0.0004	−0.043	−0.080
28. p_cause	0.076	−0.009	0.037	0.043	−0.044	−0.019	−0.071
29. p_discrep	−0.033	−0.005	0.033	−0.048	0.038	−0.086	−0.051
30. p_tentat	−0.015	0.151	0.074	−0.132	0.088	0.077	−0.096
31. p_certain	0.247**	0.004	0.184*	−0.007	0.032	0.053	−0.205**
32. p_inhib	0.063	0.179*	0.083	0.171*	−0.025	0.173*	−0.158*
33. p_incl	−0.092	0.008	0.139	−0.062	0.197*	0.028	−0.046
34. p_excl	0.044	0.091	0.159*	−0.070	0.013	0.108	−0.137
35. p_percept	−0.018	−0.013	0.070	−0.099	0.188*	0.126	0.080
36. p_see	−0.012	−0.010	0.090	−0.115	0.169*	0.115	0.082
37. p_relativ	−0.212**	−0.033	0.025	−0.176*	−0.061	−0.149	0.032
38. p_motion	−0.021	−0.011	−0.053	0.040	−0.071	−0.081	−0.100
39. p_space	−0.118	−0.012	0.190*	−0.235**	0.137	0.026	−0.031
40. p_time	−0.099	0.040	−0.144	0.013	−0.157*	−0.118	0.125
41. p_work	−0.009	−0.062	0.354***	−0.321***	0.052	0.029	−0.295***
42. p_achieve	−0.132	0.058	0.016	−0.168*	0.086	0.103	−0.139

Table F.22
Spearman's correlation of semantic variables in interactions in the direct condition (Part 3).

Variable	n_see	n_relativ	n_motion	n_space	n_time	n_work	n_achieve
22. p_social	−0.066	−0.173*	0.037	−0.219**	0.116	−0.117	−0.072
23. p_affect	0.117	−0.063	0.052	−0.191*	0.162*	−0.097	−0.060
24. p_posemo	0.107	−0.067	0.055	−0.192*	0.163*	−0.101	−0.060
25. p_negemo	0.055	0.019	−0.074	0.014	−0.041	0.086	0.024
26. p_cogmech	−0.088	−0.025	0.109	0.009	−0.003	−0.083	−0.074
27. p_insight	−0.061	−0.149	0.035	−0.115	0.012	−0.062	−0.162*
28. p_cause	−0.068	0.160*	0.094	0.187*	−0.013	0.020	0.125
29. p_discrep	−0.063	−0.116	−0.035	0.010	−0.212**	0.059	−0.177*
30. p_tentat	−0.082	0.038	−0.103	0.116	−0.064	0.076	0.086
31. p_certain	−0.202*	−0.113	−0.025	−0.064	−0.071	0.012	0.180*
32. p_inhib	−0.143	−0.008	0.139	−0.011	−0.028	−0.027	0.036
33. p_incl	−0.059	−0.076	−0.099	−0.026	−0.134	0.139	−0.171*
34. p_excl	−0.127	0.041	0.004	0.042	0.066	−0.028	0.119
35. p_percept	0.073	−0.177*	−0.153	−0.079	−0.130	0.047	−0.215**
36. p_see	0.074	−0.204*	−0.164*	−0.102	−0.143	0.054	−0.209**
37. p_relativ	0.044	0.212**	−0.027	0.213**	−0.020	0.140	0.096
38. p_motion	−0.091	0.080	0.124	0.129	−0.055	−0.003	−0.012
39. p_space	−0.029	0.072	−0.190*	0.147	−0.059	0.104	0.136
40. p_time	0.114	0.114	0.106	0.012	0.062	0.052	−0.042
41. p_work	−0.285***	0.023	−0.196*	0.113	−0.267***	0.564***	0.027
42. p_achieve	−0.173*	−0.032	−0.208**	0.049	−0.124	0.231**	−0.127

References

Araujo, T., & Bol, N. (2024). From speaking like a person to being personal: The effects of personalized, regular interactions with conversational agents. *Computers in Human Behavior: Artificial Humans*, *2*(1), Article 100030. http://dx.doi.org/10.1016/j.chbah.2023.100030, URL http://dx.doi.org/10.1016/j.chbah.2023.100030.

Bell, L. (2003). *Linguistic adaptations in spoken human-computer dialogues-empirical studies of user behavior* (Ph.D. thesis), Institutionen för talöverföring och musikakustik.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *57*(1), 289–300.

Bhattacharjee, D. R., Kuanr, A., Pradhan, D., & Moharana, T. R. (2025). Fun or warm: How conversational style boosts customer engagement. *Journal of Retailing and Consumer Services*, *85*, Article 104293.

Bickmore, T. W., & Picard, R. W. (2005). Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *12*(2), 293–327.

Brennan, S. E. (1990). Conversation as direct manipulation: An iconoclastic view. *The Art of Human-Computer Interface Design*, 393–404.

Brown, P., & Levinson, S. C. (1987). *vol. 4*, *Politeness: Some universals in language usage*. Cambridge University Press.

Brunyé, T. T., Wood, M. D., Houck, L. A., & Taylor, H. A. (2017). The path more travelled: Time pressure increases reliance on familiar route-based strategies during navigation. *Quarterly Journal of Experimental Psychology*, *70*(8), 1439–1452. http://dx.doi.org/10.1080/17470218.2016.1187637.

Burgoon, J. K., & Hubbard, A. E. (2005). Cross-cultural and intercultural applications of expectancy violations theory and interaction adaptation theory. In *Theorizing about intercultural communication* (pp. 149–171). CA: Sage Thousand Oaks.

Cai, N., Gao, S., & Yan, J. (2024). How the communication style of chatbots influences consumers' satisfaction, trust, and engagement in the context of service failure. *Humanities and Social Sciences Communications*, *11*(1), 1–11.

Chowa, S. S., Alvi, R., Rahman, S. S., Rahman, M. A., Raiaan, M. A. K., Islam, M. R., Hussain, M., & Azam, S. (2025). From language to action: A review of large language models as autonomous agents and tool users. http://dx.doi.org/10.48550/ARXIV.2508.17281, URL https://arxiv.org/abs/2508.17281.

Coluccia, E., & Louse, G. (2004). Gender differences in spatial orientation: A review. *Journal of Environmental Psychology*, *24*(3), 329–340.

Cross, S. E., & Madson, L. (1997). Models of the self: self-construals and gender. *Psychological Bulletin*, *122*(1), 5.

Dahmani, L., Idriss, M., Konishi, K., West, G. L., & Bohbot, V. D. (2023). Sex differences in spatial tasks: Considering environmental factors, navigation strategies, and age. *Frontiers in Virtual Reality*, *4*, Article 1166364. http://dx.doi.org/10.3389/frvir.2023.1166364.

Deng, Z., & Yan, J. (2025). The effect of perceived warmth, competence, and social presence of AI-driven chabots on consumers' engagement and satisfaction. *SAGE Open*, *15*(3), Article 21582440251365438.

Diederich, S., Brendel, A. B., Morana, S., & Kolbe, L. (2022). On the design of and interaction with conversational agents: An organizing and assessing review of human-computer interaction research. *Journal of the Association for Information Systems*, *23*(1), 96–138.

Ding, Y., & Najaf, M. (2024). Interactivity, humanness, and trust: A psychological approach to AI chatbot adoption in e-commerce. *BMC Psychology*, *12*(1), 595.

D'Mello, S., Dieterle, E., & Duckworth, A. (2017). Advanced, analytic, automated (AAA) measurement of engagement during learning. *Educational Psychologist*, *52*(2), 104–123.

Feine, J., Gnewuch, U., Morana, S., & Maedche, A. (2019). A taxonomy of social cues for conversational agents. *International Journal of Human-Computer Studies*, *132*, 138–161. http://dx.doi.org/10.1016/j.ijhcs.2019.07.009.

Fiske, S. T., Cuddy, A. J., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, *11*(2), 77–83. http://dx.doi.org/10.1016/j.tics.2006.11.005.

Fiske, S. T., Cuddy, A. J., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*(6), 878–902. http://dx.doi.org/10.1037/0022-3514.82.6.878.

Gefen, D., & Straub, D. W. (1997). Gender differences in the perception and use of e-mail: An extension to the technology acceptance model. *MIS Quarterly*, 389–400.

Giles, H., Edwards, A. L., & Walther, J. B. (2023). Communication accommodation theory: Past accomplishments, current trends, and future prospects. *Language Sciences*, *101*, Article 101571. http://dx.doi.org/10.1016/j.langsci.2023.101571.

Giles, H., & Ogay, T. (2013). Communication accommodation theory. In *Explaining communication* (pp. 325–344). Routledge.

Gnewuch, U., Morana, S., Adam, M., & Maedche, A. (2018). Faster is not always better: Understanding the effect of dynamic response delays in human-chatbot interaction. In *Research papers, 26th European conference on information systems* (p. 113). URL https://aisel.aisnet.org/ecis2018_rp/113.

Go, E., & Sundar, S. S. (2019). Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions. *Computers in Human Behavior*, *97*, 304–316. http://dx.doi.org/10.1016/j.chb.2019.01.020.

Gray, J., & Laidlaw, H. (2004). Improving the measurement of communication satisfaction. *Management Communication Quarterly*, *17*(3), 425–448. http://dx.doi.org/10.1177/0893318903257980.

Hecht, M. L. (1978). The conceptualization and measurement of interpersonal communication satisfaction. *Human Communication Research*, *4*(3), 253–264.

Hofstede, G. (1984). *vol. 5*, *Culture's consequences: International differences in work-related values*. sage.

Islind, A. S., Óskarsdóttir, M., Smith, S. Á., & Arnardóttir, E. S. (2023). The friendly chatbot: Revealing why people use chatbots through a study of user experience of conversational agents. In *14th Scandinavian conference on information systems*. URL https://aisel.aisnet.org/scis2023/6.

Kulms, P., & Kopp, S. (2018). A social cognition perspective on human–computer trust: The effect of perceived warmth and competence on trust in decision-making with computers. *Frontiers in Digital Humanities*, *5*, http://dx.doi.org/10.3389/fdigh.2018.00014.

Lakey, B., & Cohen, S. (2000). Social support theory and measurement. *Social Support Measurement and Intervention: A Guide for Health and Social Scientists*, *2952*.

Lawton, C. A. (1994). Gender differences in way-finding strategies: Relationship to spatial ability and spatial anxiety. *Sex Roles*, *30*(11), 765–779.

Lee, M. K., Kiesler, S., & Forlizzi, J. (2011). Mining behavioral economics to design persuasive technology for healthy choices. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 325–334). ACM, http://dx.doi.org/10.1145/1978942.1978989.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *46*(1), 50–80. http://dx.doi.org/10.1518/hfes.46.1.50_30392.

Liebrecht, C., Sander, L., & Van Hooijdonk, C. (2020). Too informal? How a chatbot's communication style affects brand attitude and quality of interaction. In *International workshop on chatbot research and design* (pp. 16–31). Springer.

Luger, E., & Sellen, A. (2016). " like having a really bad PA" The Gulf between user expectation and experience of conversational agents. In *Proceedings of the 2016 CHI conference on human factors in computing systems* (pp. 5286–5297).

Mairesse, F., & Walker, M. (2011). Controlling user perceptions of linguistic style: Trainable generation of personality traits. *Computational Linguistics*, *37*(3), 455–488.

Mendes, G. A., & Martins, B. (2023). Quantifying valence and arousal in text with multilingual pre-trained transformers. In *European conference on information retrieval* (pp. 84–100). Springer.

Morozov, E. (2013). *To save everything, click here: The folly of technological solutionism*. PublicAffairs.

Nass, C., Fogg, B., & Moon, Y. (1996). Can computers be teammates? *International Journal of Human-Computer Studies*, *45*(6), 669–678. http://dx.doi.org/10.1006/ijhc.1996.0073.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, *56*(1), 81–103.

Nazareth, A., Huang, X., Voyer, D., & Newcombe, N. (2019). A meta-analysis of sex differences in human navigation skills. *Psychonomic Bulletin & Review*, *26*(5), 1503–1528.

Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *52*(3), 381–410. http://dx.doi.org/10.1177/0018720810376055.

Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology*, *77*(6), 1296.

Poivet, R., Lopez Malet, M., Pelachaud, C., & Auvray, M. (2023). The influence of conversational agents' role and communication style on user experience. *Frontiers in Psychology*, *14*, Article 1266186. http://dx.doi.org/10.3389/fpsyg.2023.1266186.

Pralat, N., Ischen, C., & Voorveld, H. (2024). Feeling understood by AI: How empathy shapes trust and influences patronage intentions in conversational AI. In *International symposium on chatbots and human-centered AI* (pp. 234–259). Springer.

Qin, J., Nan, Y., Li, Z., & Meng, J. (2025). Effectiveness of communication competence in AI conversational agents for health: Systematic review and meta-analysis. *Journal of Medical Internet Research*, *27*, Article e76296.

Reeves, B., & Nass, C. (1996). *vol. 10*, *The media equation: How people treat computers, television, and new media like real people* (pp. 19–36). Cambridge, UK: (10).

Rheu, M., Dai, Y., Meng, J., & Peng, W. (2024). When a chatbot disappoints you: Expectancy violation in human-chatbot interaction in a social support context. *Communication Research*, *51*(7), 782–814.

Ribino, P. (2023). The role of politeness in human–machine interactions: A systematic literature review and future perspectives. *Artificial Intelligence Review*, *56*(Suppl 1), 445–482.

Ruijten, P. A., Midden, C. J., & Ham, J. (2015). Lonely and susceptible: The influence of social exclusion and gender on persuasion by an artificial agent. *International Journal of Human-Computer Interaction*, *31*(11), 832–842.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68.

Schinazi, V. R., Meloni, D., Grübel, J., Angus, D. J., Baumann, O., Weibel, R. P., Jeszenszky, P., Hölscher, C., & Thrash, T. (2023). Motivation moderates gender differences in navigation performance. *Scientific Reports*, *13*(15995), http://dx.doi.org/10.1038/s41598-023-43241-4.

Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction*, *36*(6), 495–504.

Skitka, L. J., Mosier, K., & Burdick, M. D. (2000). Accountability and automation bias. *International Journal of Human-Computer Studies*, *52*(4), 701–717. http://dx.doi.org/10.1006/ijhc.1999.0349.

Sobieraj, S., & Krämer, N. C. (2020). Similarities and differences between genders in the usage of computer with different levels of technological complexity. *Computers in Human Behavior*, *104*, Article 106145. http://dx.doi.org/10.1016/j.chb.2019.09.021.

Sterken, R. K., & Kirkpatrick, J. R. (2024). Conversational alignment with artificial intelligence in context. *Philosophical Perspectives*, *38*(1), 89–102. http://dx.doi.org/10.1111/phpe.12205.

Svenningsson, N., & Faraon, M. (2019). Artificial intelligence in conversational agents: A study of factors related to perceived humanness in chatbots. In *AICCC 2019*, *Proceedings of the 2019 2nd artificial intelligence and cloud computing conference* (pp. 151–161). ACM, http://dx.doi.org/10.1145/3375959.3375973.

Tannen, D. (1990). *You just don't understand: Women and men, Conversation*. New York: Ballantine Books.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, *29*(1), 24–54.

Tenenbaum, H. R., Ford, S., & Alkhedairy, B. (2011). Telling stories: Gender differences in peers' emotion talk and communication style. *British Journal of Developmental Psychology*, *29*(4), 707–721.

Wegmann, A., Schraagen, M., & Nguyen, D. (2022). Same author or just same topic? Towards content-independent style representations. In S. Gella, H. He, B. P. Majumder, B. Can, E. Giunchiglia, S. Cahyawijaya, S. Min, M. Mozes, X. L. Li, I. Augenstein, A. Rogers, K. Cho, E. Grefenstette, L. Rimell, & C. Dyer (Eds.), *Proceedings of the 7th workshop on representation learning for NLP* (pp. 249–268). Dublin, Ireland: Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/2022.repl4nlp-1.26, URL https://aclanthology.org/2022.repl4nlp-1.26/.

Yi, Z., Ouyang, J., Xu, Z., Liu, Y., Liao, T., Luo, H., & Shen, Y. (2025). A survey on recent advances in LLM-based multi-turn dialogue systems. *ACM Computing Surveys*, *58*(6), 1–38. http://dx.doi.org/10.1145/3771090.

Zellou, G., & Holliday, N. (2024). Linguistic analysis of human-computer interaction. *Frontiers in Computer Science*, *6*, Article 1384252.