Impact of Personalised AI Chat Assistant on Mediated Human-Human Textual Conversations: Exploring Female-Male Differences

Conference Paper · March 2024

DOI: 10.1145/3640544.3645218

CITATIONS

0

4 authors:

Jindi Wang

Durham University

13 PUBLICATIONS **14** CITATIONS

SEE PROFILE

Ioannis Ivrissimtzis

Durham University

115 PUBLICATIONS **823** CITATIONS

SEE PROFILE

Zhaoxing Li

University of Southampton

18 PUBLICATIONS **34** CITATIONS

SEE PROFILE

Lei Shi

Newcastle University

118 PUBLICATIONS **1,357** CITATIONS

SEE PROFILE

Impact of Personalised AI Chat Assistant on Mediated Human-Human Textual Conversations: Exploring Female-Male Differences

JINDI WANG, Durham University, UK
IOANNIS IVRISSIMTZIS, Durham University, UK
ZHAOXING LI, University of Southampton, UK
LEI SHI, Newcastle University, UK

ABSTRACT

In text-based communication, the enhancement of user experiences during the earliest stages of contact, particularly while establishing rapport, is of utmost importance. This research investigates differences between males and females in the utilisation of an AI conversation assistant at the beginning of a conversation. The system has text "Recommendation" and "Polishing" capabilities and can customise the linguistic style, selecting either a *humorous* or a *respectful* tone. Users could also choose between three different levels of AI extraversion. In user evaluation studies, the system received a favourable usability rating, confirming its efficacy. Generally, male users reported a greater level of user experience compared to female users. However, both male and female users indicated an elevated sense of comfort and a greater inclination to sustain connections while utilising the AI system.

CCS Concepts: • **Human-centered computing → Empirical studies in HCI**.

Additional Key Words and Phrases: Chat assistant, User study, Personalisation, Conversational AI, Computer-Assisted Human Interaction

ACM Reference Format:

Jindi Wang, Ioannis Ivrissimtzis, Zhaoxing Li, and Lei Shi. 2024. Impact of Personalised AI Chat Assistant on Mediated Human-Human Textual Conversations: Exploring Female-Male Differences. In *29th International Conference on Intelligent User Interfaces - Companion (IUI Companion '24), March 18–21, 2024, Greenville, SC, USA.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3640544.3645218

1 INTRODUCTION

Recent advancements in language models like GPT-4 [10] allow artificial intelligence (AI) to analyse sentiment [14] and then mimic specific writing styles [7]. The rise of AI-powered chat assistants signifies a transformative shift in computer-assisted human contact, not only enhancing efficiency but also impacting fundamental elements of human connection [12].

Crucial to both digital and face-to-face interactions is the initial encounter, often termed "icebreaking". Despite its brevity, this stage significantly influences subsequent interactions in professional and informal settings [13]. Integrating AI chat assistants in these early phases, with features like text recommendation and polishing, can shape the user experience. While researchers have extensively explored the utility of Conversational AI in task-oriented scenarios [3], limited attention has been given to their potential to improve human-human dialogues. This study focuses on personalised AI chat assistants, with text recommendation and polishing functionalities, in human-human initial conversations.

Manuscript submitted to ACM

We investigate potential male and female differences in user experience, an area with limited literature [6, 9]. Recognising that gender influences communication styles, expectations, and perceptions, understanding how males and females interact with AI assistants can inform the design of more personalised and effective systems [4, 8], providing insights into how AI chat assistants deployed on specific application domains can cater to diverse user needs, including male-female differences [15].

In summary, acknowledging the significant influence of icebreaking on the development of following interactions and seeing the absence of any existing chat application that includes a customised AI assistant, this research centres on this particular domain of application, seeking to answer "***How do user experiences differ between males and females when utilising a personalised chat assistant in the context of initial text conversations?***" Our aim is to obtain insights that would influence the design of the next generation of AI chat assistants, making them more user-centric and context-aware.

2 DESIGN OF PERSONALISED CHAT ASSISTANTS

2.1 Preliminary User Survey

We conducted a preliminary user survey to gauge interest and views on the necessity for personalisation among prospective users. The participants of the preliminary survey were briefed on the study's aims and methodologies, and consent was obtained following our institution's guidelines. An online questionnaire[1] was administered to a diverse sample of 162 participants aged 15 to 44. The survey explored attitudes towards chatbots and AI helpers, current satisfaction levels, and areas for improvement.

Participants rated various language styles for their ideal AI chat assistant using a scale from 1 ("not needed at all") to 5 ("highly needed"). In the analysis of the results, scores exceeding 3 were considered indicative of favourable sentiments. A majority (82.67%) desired a more personalised AI chat experience. "Respectful" and "Humorous" styles were most favoured (67.72% and 74.16%, respectively). "Playful" and "Serious" styles were preferred by 66.93% and 53.54%, respectively. A minority (33.86%) preferred an "Offensive" tone.

Acknowledging user preferences for personalisation and their favoured language styles ("Respectful" and "Humor"), subsequent design prioritised these styles. Our approach anchored our research in authentic user preferences, enhancing its relevance and practicality.

2.2 Key Features of Personalised Chat Assistant

The user interface incorporates an AI chat assistant specifically developed to enhance the overall communication experiences of users. The diagram in Fig. 1 illustrates the two main functionalities of the system.

Text Recommendation: This functionality proved to be advantageous in situations where the users were uncertain regarding appropriate responses inside conversational contexts. The AI provides a recommended response at a pre-established AI extraversion level and language style, based on the user's preferences:

- **Extraversion Level**: Three distinct levels were supported: introverted, average, and extroverted. This allows the AI assistant's suggested responses to align with the user's preferred mode of interaction.
- **Linguistic Style**: Individuals can select between the options of "humorous" and "respectful". This feature provides a second customisation level, catering to the user's preferences and their perception of the conversation's context.

[1] https://qfreeaccountssjc1.az1.qualtrics.com/jfe/form/SV_0VQb8OE6ZM5R5u6

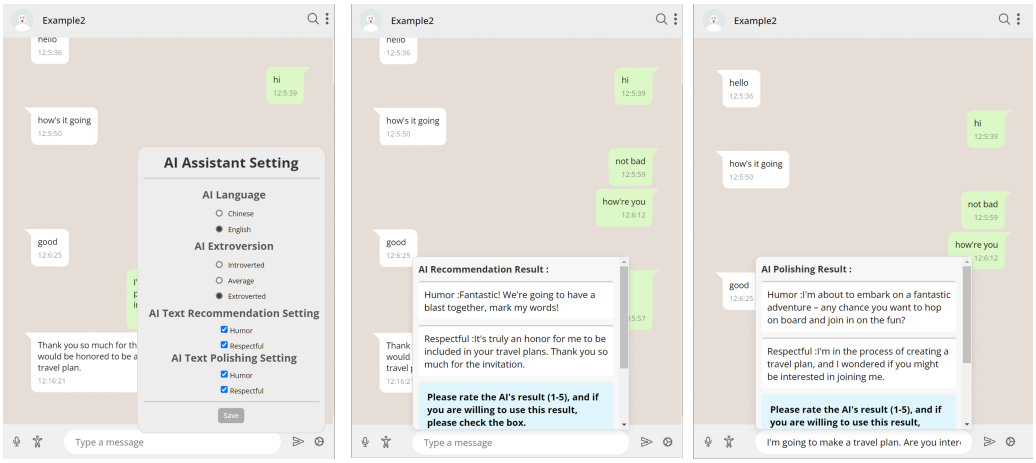


Fig. 1. Sample interactions between users and AI chat assistants are shown in English; however, all participants in the study consented to converse in Chinese.

Text Polishing: In addition to providing suggestions for responses, the assistant can enhance and improve messages written by the user. Text polishing is tailored to the user's preferences by choices of extraversion level and linguistic style, with exactly the same options as the recommendation functionality. The symmetry of the design enhances consistency, coherence, and clarity.

2.3 Implementation

The developed Personalised AI Chat Assistant, is supported by OpenAI's GPT-3.5 API[2], and provides a personalised conversational interface, enabling users to adjust the AI's amount of extraversion and linguistic style according to their preferences. The workflow of the system combines user inputs with chosen AI-enhanced features to produce improved textual outputs. Furthermore, it examines conversation data gathered from interactions with friends, meticulously analyzing the language, tone, and context and incorporates user-defined AI recommendation functionalities to produce contextually appropriate text recommendations. The system's versatility, along with a user interface influenced by the design of WhatsApp's web platform and supported by an SQLite database for effective data storage, tailored for analyzing user-focused digital communication, is detailed in an Appendix with a high-level flowchart.

3 USER STUDY

3.1 Participants & Study Setup

To obtain insights into user interactions with our Personalised AI Chat Assistant, we experimented with a cohort of 18 Chinese participants. Following a comprehensive explanation of the objectives of the study, the participants signed consent forms, indicating their voluntary agreement to partake in the experimental procedures. Gender and age information for each participant can be found in the Appendix. The participants' briefing was that they would chat with unfamiliar people and that they should mainly focus on the topic of travel. This is to ensure our experiment was conducted on the same conditions and conversation topic.

[2]https://platform.openai.com/docs/api-reference

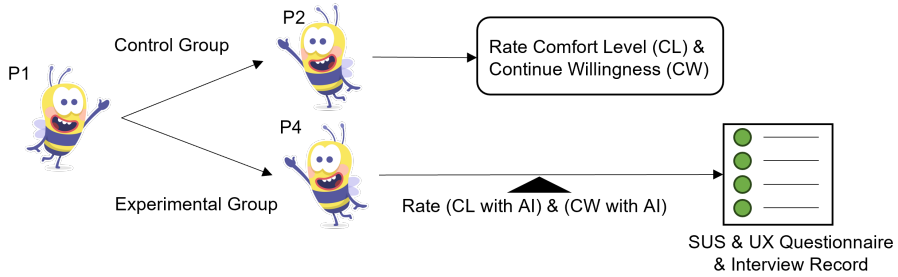


Fig. 2. Control group for an initial chat without AI followed by AI-assisted chat, with Comfort Level (CL), Continuation Willingness (CW) assessments, and experimental group for post-experiment user experience questionnaire and interview.

Participants Pairing: We employed a random pairing approach with two restrictions: all pairs consisted of one male and one female participant, and secondly, each participant engaged with different persons in their AI-assisted and non-assisted conversations.

Experimental Process: Each participant engaged in two separate sessions, as illustrated in Fig. 2. **1) The control session** - a text conversation with someone they had no prior acquaintance with, without using the assistance of the AI chatbot. After the interaction, participants evaluated their comfort levels during the chat and their inclination to sustain the dialogue. **2) The experimental session** - a subsequent text conversation with a new acquaintance, distinct from that of the control session, using this time the functionalities of the AI chat assistant. Throughout this session, participants received AI-generated *recommendations* and *polishing*. To ensure consistency, participants adhered to two predetermined linguistic styles while having the freedom to select from various AI extraversion levels, and each session was carefully regulated to last approximately 20 minutes.

At the end of this session, participants were invited to fill out the System Usability Scale (SUS) and User Experience (UX) questionnaire[3], a task lasting about 5 minutes.

3.2 Metrics

System Usability Scale (SUS): A validated questionnaire assessing user perceptions of usability in personalised AI chat assistants. It gauges user-friendliness and comprehensibility, with higher scores indicating a more user-friendly interface and a sense of control for users.

User Experience (UX): A metric evaluating the overall user experience during interactions with the AI chat assistant. It includes a 7-point Likert scale of efficiency, effectiveness, and user satisfaction. A positive UX score indicates that the AI assistant meets or exceeds user expectations in terms of usefulness and response [11].

Comfort Level (CL): A metric measuring user comfort during initial textual interactions on online platforms. Rated on a scale from 1 to 5, it reflects ease or tension in conversation, providing insights for optimising chat interfaces and interaction experiences [2].

Continue Willingness (CW): A metric assessing users' desire to extend textual interactions after initial conversations. Scored from 1 to 5, it helps understand engagement dynamics, emphasizing the importance of fostering continued conversations for deeper connections [5].

[3]https://qfreeaccountssjc1.az1.qualtrics.com/jfe/form/SV_3acj95ynXNOclgO

4 RESULTS

4.1 Usability and User Experience

First, we utilised the System Usability Scale (SUS) as a means of measurement of the perceived usability of the system. With average System Usability Scale (SUS) scores of 70.83 (SD = 3.73) for males and 70.94 (SD = 6.37) for females, both male and female users perceived the usability of the system as "good" [1], with minimal, statistically insignificant differences between them.

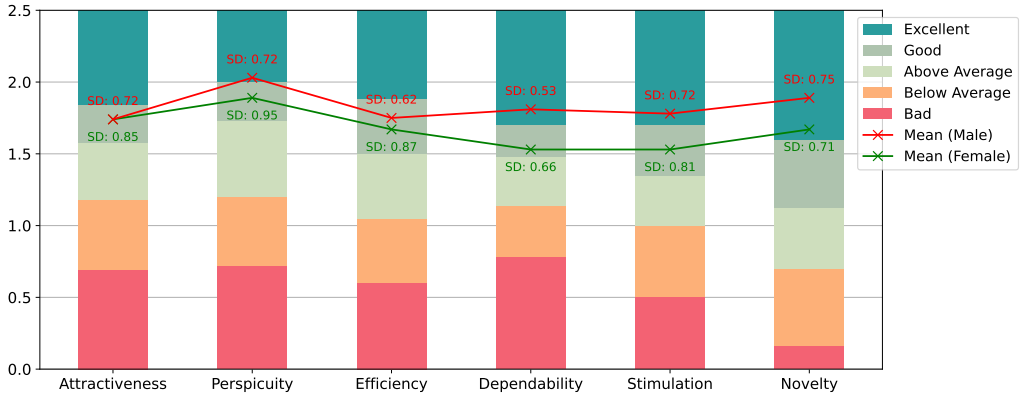


Fig. 3. Distribution of scores across six key dimensions of male and female user experience.

Table 1. Cronbachs Alpha-Coefficient for male and females.

Scales	Male	Female
Attractiveness	0.97	0.92
Perspicuity	0.96	0.98
Efficiency	0.95	0.89
Dependability	0.82	0.88
Stimulation	0.90	0.92
Novelty	0.97	0.94

Using the six UX dimensions outlined by Schrepp *et al.* [11], our questionnaire revealed the system's strong performance. As depicted in Fig. 3, the AI chat assistant demonstrated exceptional performance in the attractiveness and efficiency dimensions, receiving a commendable grade of "Good" from both gender groups. "Novelty" also earned an "Excellent" rating for innovation and engagement from both groups. However, a difference between female and male groups emerged for perspicuity, dependability, and stimulation. While male users rated these dimensions as "Excellent", female users rated them as "Good". This indicates that the system's design and engagement features resonate well with both female and male groups, but female users would have preferred further refinement. Encouragingly, the internal consistency of our evaluation, gauged by Cronbach's Alpha, exceeded the empirical 0.7 threshold for all dimensions, as shown in Table 1. This high reliability strengthens the validity of our findings, adding confidence to the observed gender differences in UX.

4.2 Comfort Level and Continue Willingness

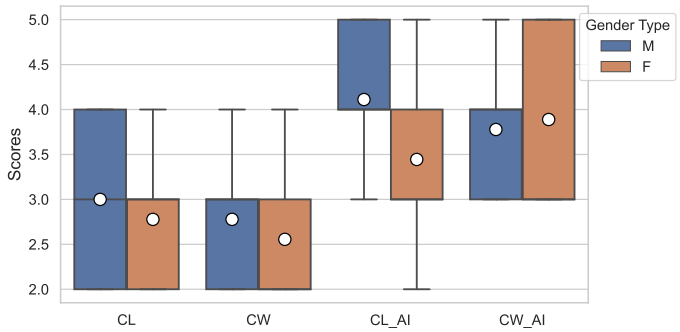


Fig. 4. Boxplots for males and females of the chat comfort level (CL) and the continue willingness (CW).

Fig. 4 shows boxplots of comfort level (CL) and continue willingness (CW) scores for male and female users.

Comfort Level without AI (CL): Male users reported a slightly higher average comfort level (M = 3.00) compared to female users (M = 2.78). While their difference is not statistically significant ($p = 0.28$), we note the wider interquartile range (spread of comfort levels) among male users, a finding that might warrant further investigation.

Continuation Willingness without AI (CW): Male users expressed a slightly higher willingness to continue the conversation (M = 2.78) compared to females (M = 2.56), which again was not statistically significant. The similar spread of responses within both groups, as evidenced by the interquartile range, suggests that individuals of both genders displayed a comparable range of engagement, with some more interested and others less enthusiastic about further conversation.

Comfort Level with AI (CL-AI): Male users reported a statistically significant ($p = 0.02$) higher level of comfort level (M = 4.11) compared to females (M = 3.44). Interestingly, the female user group displayed a wider range of comfort levels, suggesting a more varied AI experience. This could be an indication that certain aspects of the AI's assistance resonate more strongly with some female users than others, something that might warrant further investigation.

Continue Willingness with AI (CW-AI): While both genders showed a high overall willingness to continue interacting with AI, response distributions differed. Male users had a higher degree of consistency in their scores, resulting to a tighter interquartile range. Moreover, the mean score of the female user (M = 3.89) surpassed that of males (M = 3.78), even though the difference was not statistically significant, ($p = 0.41$). These findings could suggest that specific features of the AI implementation resonated very well with some of female users and stimulated their inclination to continue the interaction, while male users maintained a more consistent level of interest.

Overall, the adoption of the AI chat assistant significantly increased users' comfort levels and willingness to continue interaction ($p << 0.05$). The 37.04% and 24.00% increases in the mean values of comfort levels for males and females, respectively, and the 36.00% and 52.17% increases in the willingness to continue, indicate that AI had a positive impact on the overall user experience.

4.3 Discussion

4.3.1 ***Usability and User Experience Reflection***. The examination of the AI chat assistant's user experience and usability reveals its potential to enhance text-based initial conversation situations, especially for its intuitive

interface and ease of use. The AI chat assistant received high ratings for **Novelty** and **Perspicuity**, due to its unique conversational approach and the clarity of its interface. Nevertheless, user feedback pointed to areas for improvement. Enhancing the visual appeal through customisable themes, expanding the personalisation options, and further refining the language model's context awareness could all contribute to an even more engaging and seamless UX.

4.3.2 ***Comfort Levels and Continuation Willingness.*** Our empirical findings demonstrate a significant increase in the mean values of comfort levels and willingness to continue conversations for both males and females, highlighting the potential of AI to mediate and facilitate positive user experiences, especially in unorganised social dialogues where contextual information is sparse. Specifically, the AI's engaging prompts, playful humour, and icebreaker questions seem to contribute to a favourable atmosphere, highlighting its ability to create a favourable atmosphere for ongoing interactions, even during the early phases of the unorganised dialogue, where there is less contextual information available to guide the language model. On the other hand, it's important to acknowledge that AI's effectiveness can vary depending on the specific topic and the quality of its training data, highlighting the need for further research and development in this domain.

4.3.3 ***Limitations and Future Directions.*** While demonstarting the potential of artificial intelligence in facilitating initial text interactions, particularly within the Chinese population segment studied, the present study also highlights the need for greater diversity in future studies. Moreover, the primary focus on specific characteristics such as gender, further limits the generalisability of the findings. To ensure broader applicability, future research should incorporate a wider range of subjects in terms of age, ethnicity, geographical location, and socio-economic background, allowing for a more nuanced understanding of how AI chat assistants can facilitate conversations within diverse populations.

It is worth noting that even subtle variations in AI behaviour within the same extraversion level can significantly influence human impressions, highlighting the need for precise tuning of AI communication styles. Moreover, our study, and several others on the topic, suggest a trend towards over-reliance on the text polishing feature, potentially raising concerns about user agency and a long-term impact on communication skills. Investigating this phenomenon and its long-term effects on genuine self-expression and language development, should be a priority for future research.

Furthermore, the study revealed an intrinsic subjectivity in the experimental design, suggesting that users might interpret the AI responses in diverse ways. This highlights the need for future research to account for user variability. One additional limitation is that using predetermined AI personality profiles limits adaptability to individuals with diverse communication styles. Future studies should explore AI models that can dynamically adjust their personalities to better cater to user preferences. Finally, this study did not assess the nuanced differences between various AI features such as text *recommendation* and *polishing*, leaving a promising research direction for further investigations.

5 CONCLUSION

This digital age revolves around text-based interactions, making the initial texting experience crucial for building positive communication dynamics. Our findings highlight the significant potential of AI interventions to enhance user experience in these initial encounters. By fostering comfort and encouraging sustained conversations, AI chat assistants can bridge the gap and establish meaningful connections. However, we must acknowledge the inherent limitations of such innovations, and particularly the risk of over-reliance on the polishing feature and its potential impact on the users' communication skills. Additionally, ensuring consistent AI behaviour and adapting to the diverse personalities of real-world users pose crucial challenges for future iterations.

REFERENCES

[1] Aaron Bangor, Philip Kortum, and James Miller. 2009. Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of usability studies* 4, 3 (2009), 114–123.

[2] Susan Bergin and Ronan Reilly. 2005. The influence of motivation and comfort-level on learning to program. In *Proc. of the 17th Workshop of the Psychology of Programming Interest Group, PPIG 05*. Psychology of Programming Interest Group, 293–304.

[3] Hongshen Chen, Xiaorui Liu, Dawei Yin, and Jiliang Tang. 2017. A survey on dialogue systems: Recent advances and new frontiers. *Acm Sigkdd Explorations Newsletter* 19, 2 (2017), 25–35.

[4] Jasper Feine, Ulrich Gnewuch, Stefan Morana, and Alexander Maedche. 2020. Gender bias in chatbot design. In *Chatbot Research and Design: Third International Workshop, CONVERSATIONS 2019, Amsterdam, The Netherlands, November 19–20, 2019, Revised Selected Papers 3*. Springer, 79–93.

[5] Mark Freiermuth and Douglas Jarrell. 2006. Willingness to communicate: can online chat help? 1. *International journal of applied linguistics* 16, 2 (2006), 189–212.

[6] Yiting Guo, De Liu, Ximing Yin, and Sean Xin Xu. 2020. "She is not just a computer"': Gender role of AI chatbots in debt collection. In *International Conference on Information Systems-Making Digital Inclusive: Blending the Local and the Global, ICIS 2020*. Association for Information Systems.

[7] Zhiting Hu, Zichao Yang, Xiaodan Liang, Ruslan Salakhutdinov, and Eric P Xing. 2017. Toward controlled generation of text. In *International Conference on Machine Learning*. PMLR, 1587–1596.

[8] Marian McDonnell and David Baxter. 2019. Chatbots and gender stereotyping. *Interacting with Computers* 31, 2 (2019), 116–121.

[9] Daniela SM Pereira, Filipe Falcão, Lilian Costa, Brian S Lunn, José Miguel Pêgo, and Patrício Costa. 2023. Here's to the future: Conversational agents in higher education-a scoping review. *International Journal of Educational Research* 122 (2023), 102233.

[10] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *The Journal of Machine Learning Research* 21, 1 (2020), 5485–5551.

[11] Martin Schrepp, Jorg Thomaschewski, and Andreas Hinderks. 2017. Construction of a Benchmark for the User Experience Questionnaire (UEQ). *International Journal of Interactive Multimedia and Artificial Intelligence* 4, 4 (2017), 40–44.

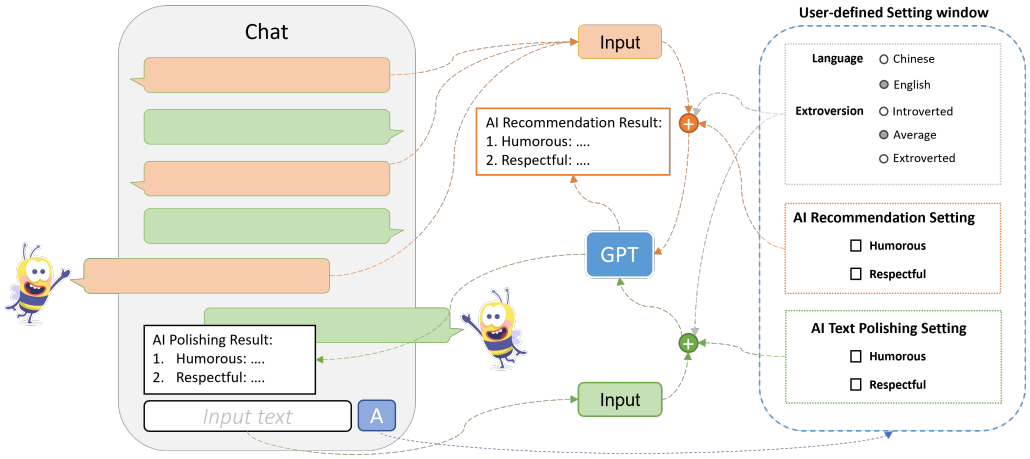
[12] Iulian V Serban, Chinnadhurai Sankar, Mathieu Germain, Saizheng Zhang, Zhouhan Lin, Sandeep Subramanian, Taesup Kim, Michael Pieper, Sarath Chandar, Nan Rosemary Ke, et al. 2017. A deep reinforcement learning chatbot. *arXiv:1709.02349* (2017).

[13] Bayan Abu Shawar and Eric Atwell. 2007. Chatbots: are they really useful? *Journal for Language Technology and Computational Linguistics* 22, 1 (2007), 29–49.

[14] G Vinodhini and RM Chandrasekaran. 2012. Sentiment analysis and opinion mining: a survey. *International Journal* 2, 6 (2012), 282–292.

[15] Banghui Zhang, Yonghan Zhu, Jie Deng, Weiwei Zheng, Yang Liu, Chunshun Wang, and Rongcan Zeng. 2023. "I Am Here to Assist Your Tourism": Predicting Continuance Intention to Use AI-based Chatbots for Tourism. Does Gender Really Matter? *International Journal of Human–Computer Interaction* 39, 9 (2023), 1887–1903.

A SUPPLEMENTARY MATERIAL



Workflow overview of the personalised AI chat assistant powered by GPT-3.5.

Demographics and control and experimental session pairs.

Control Session Without AI (Age, Male/Female)	Experimental Session With AI
P1 (28, Male) & P2 (27, Female)	P1 & P4
P3 (28, Male) & P4 (33, Female)	P3 & P6
P5 (28, Male) & P6 (25, Female)	P5 & P8
P7 (27, Male) & P8 (27, Female)	P7 & P10
P9 (28, Male) & P10 (28, Female)	P9 & P12
P11 (27, Male) & P12 (21, Female)	P11 & P14
P13 (27, Male) & P14 (29, Female)	P13 & P16
P15 (27, Male) & P16 (35, Female)	P15 & P18
P17 (27, Male) & P18 (31, Female)	P17 & P2