USER MODELING CHALLENGES IN INTERACTIVE AI ASSISTANT SYSTEMS

Megan Su, Yuwei Bao
University of Michigan
{meganysu, yuweibao}@umich.edu

ABSTRACT

Interactive Artificial Intelligent (AI) assistant systems are designed to offer timely guidance to help human users to complete a variety tasks. One of the remaining challenges is to understand user's mental states during the task for more personalized guidance. In this work, we analyze users' mental states during task executions and investigate the capabilities and challenges for large language models to interpret user profiles for more personalized user guidance.

1 INTRODUCTION

In the digital age, there is immense potential for artificial intelligent (AI) assistant to guides users through complex tasks, from changing laptop batteries to piping frosting on a cake. One of the main challenges, however, lies in creating an interactive system that can not only understand which step the user is at, but can also detect user's mental states, such as frustration, familiarity with the task, detail-orientation, etc. Only when the model understands user's personalized preferences can it provide a user-friendly experience and promoting better task completion outcome.

In this work, we address this challenge by 1) extending the WTaG (Bao et al. (2023)) dataset to incorporate user mental profiles during the task, such as frustration, eagerness to speak, and detail-orientation, and 2) by investigating if existing large language models (LLM) can accurately interpret underlying users states. This work aims to enhance user experience with AI assistance for the next steps by tailoring responses to best accommodate user's emotional and mental dispositions.

2 DATASET - USER MODELING ANALYSIS

AI Assistant systems such as HoloAssist (Wang et al. (2023)), Ego How-TO (Xue & Grauman (2023)), Google Assistant (Google (2023)), Siri (Apple (2023)), and Alexa (Amazon (2023)) are designed to guide human users on various tasks. Some of which leverage augmented reality headsets to enable visual and verbal communications. In this work, we expanded the human instructor-user dialog from the WTaG (Bao et al. (2023)) dataset to include 6 categories of user profiles for each recording: **Frustration**, **Eagerness to ask questions**, **Talkative**, **Experience**, **Familiarity with tools**, and **Detail-orientation**. For each of the 55 recordings (across 17 users and 3 recipes), we annotated "yes" or "no" to each category above. A total of 330 data points were collected.

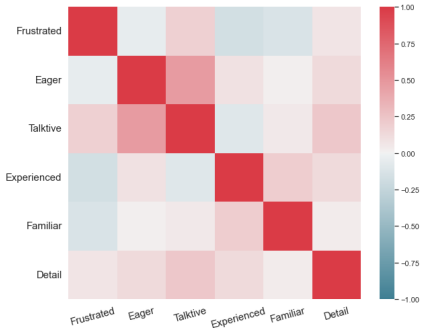


Figure 1: Feature Correlations

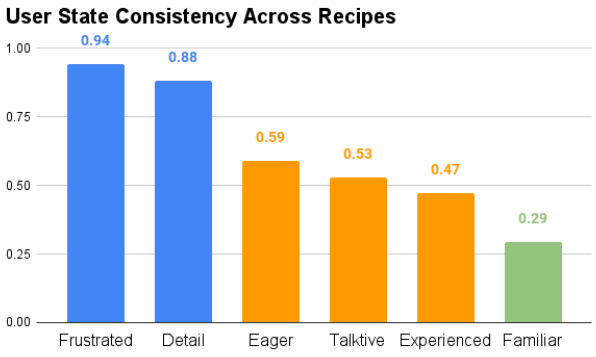
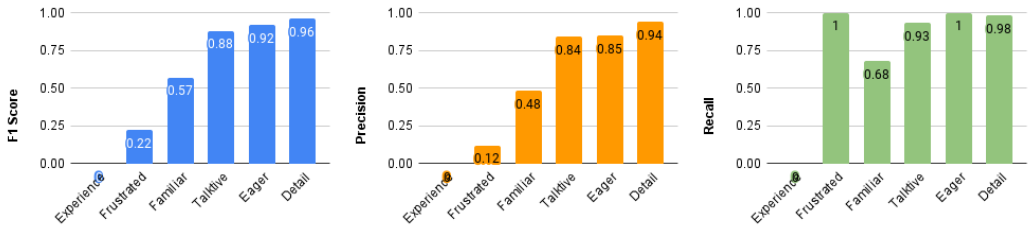


Figure 2: User Mental State Consistency Across Recipes



(a) F1 Score (b) Precision (c) Recall

Figure 3: ChatGPT User Modeling Prediction Performance

While choosing the 6 user profile categories, we try to cover a wide range of aspects to model users' mental states. We conducted a pair-wise Pearson correlation analysis to see how different features are related. Figure 1 reveals that other than "Eager to ask questions" and "Talkative", which are moderately correlated, most of the features depict a unique perspective of the user states. Other interesting observations include that users' frustration has a mild positive correlation with talkativeness and negative correlation with their experience, which suggests AI assistants could adjust their guidance according to user's proactive engagement or task-related experience level.

We also examined the user's mental state consistency across different recipe executions. For each user, the consistency is 1 if the user exhibited the same "Yes" or "No" to each profile category for all three recipes, and 0 otherwise. Among the results in Figure 2, a high consistency was observed on users for both "Frustration" and "Detail-orientation", suggesting that AI assistants could transfer their user-specific accommodations across different tasks according to user's traits on attention to details, or external factors, such as user's mood for that particular day. "Eager to ask questions", "Talkativeness" and "Experienced" are moderately consistent for each user across recipes, suggesting a combination of both user-specific accommodation and recipe-dependent adjustment. "Familiarity with tools" have the lowest consistency, which indicates the AI assistant would need to understand user's proficiency on each tool used in different tasks for more situated guidance.

3 User Modeling with Large Language Models

Several previous works (Li & Zhao (2020); Webb et al. (2001); Çelikok et al. (2023)) proposed using deep learning methods for user modeling. In this work, we investigate how large language models can interpret user's mental states during the tasks. Given the dialog history between the user and the instructor for each recording, we prompt ChatGPT[1] to predict "yes" or "no" for each user profiling category. By comparing the F1 score, precision and recall of model's predictions with our annotations, we gain insights into the model's ability to interpret users' needs and mental states. We set the temperature to zero, repeated the experiments 3 times and observed negligible variance.

Our results in Figure 3 indicate LLM's high ability to detect when users are "Detail-oriented", "Eager to ask questions", or "Talkative", with F1 scores of 0.96, 0.92, and 0.88, respectively. However, the "Frustrated" category reveals a stark contrast, in which it frequently misclassifies non-frustrated behavior, yielding a high false positive rate in this category. With a precision and recall of zero, the "Experienced" attribute needs significant improvement, pointing to necessary steps to improve models' ability to understand user's knowledge and experience on the tasks.

4 Conclusion

In this work, we analyze human users' mental states during task executions, and study how well can LLM based interactive AI assistant systems understand and accommodate for these personalized needs. We extended the WTaG dataset to incorporate 6 categories of user's mental states based on the dialog history, and challenge the LLMs to predict user states. Our analysis showed that users exhibited different levels of consistency across user profile categories and that significant improvements are needed for LLMs to understand several categories. Future work could include larger scale dataset analysis (e.g. Wang et al. (2023); Grauman et al. (2023)), and investigate better prompting strategies, online user state detection and in-context adjustments, categorized fine-tuning, and multimodal signals for better user modeling.

[1] `https://openai.com/blog/chatgpt`; Version GPT3.5-turbo-0301.

ACKNOWLEDGEMENTS

This work is supported by DARPA PTG program HR00112220003. We would like to thank Prof. Joyce Chai for her guidance and suggestions.

URM STATEMENT

The authors acknowledge that at least one key author of this work meets the URM criteria of ICLR 2024 Tiny Papers Track. The first author for this submission is participating as the lead author for the first time, and this distinction is particularly significant as all authors do not identify as cis-male and are non-White individuals – demographics that remain underrepresented in the field of AI research.

REFERENCES

Amazon. Alexa, 2023. URL https://alexa.amazon.com/.

Apple. Siri, 2023. URL https://www.apple.com/siri/.

Yuwei Bao, Keunwoo Peter Yu, Yichi Zhang, Shane Storks, Itamar Bar-Yossef, Alexander De La Iglesia, Megan Su, Xiao Lin Zheng, and Joyce Chai. Can foundation models watch, talk and guide you step by step to make a cake?, 2023.

Google. Google assistant, 2023. URL https://assistant.google.com/.

Kristen Grauman, Andrew Westbury, Lorenzo Torresani, Kris Kitani, Jitendra Malik, Triantafyllos Afouras, Kumar Ashutosh, Vijay Baiyya, Siddhant Bansal, Bikram Boote, Eugene Byrne, Zach Chavis, Joya Chen, Feng Cheng, Fu-Jen Chu, Sean Crane, Avijit Dasgupta, Jing Dong, Maria Escobar, Cristhian Forigua, Abrham Gebreselasie, Sanjay Haresh, Jing Huang, Md Mohaiminul Islam, Suyog Jain, Rawal Khirodkar, Devansh Kukreja, Kevin J Liang, Jia-Wei Liu, Sagnik Majumder, Yongsen Mao, Miguel Martin, Effrosyni Mavroudi, Tushar Nagarajan, Francesco Ragusa, Santhosh Kumar Ramakrishnan, Luigi Seminara, Arjun Somayazulu, Yale Song, Shan Su, Zihui Xue, Edward Zhang, Jinxu Zhang, Angela Castillo, Changan Chen, Xinzhu Fu, Ryosuke Furuta, Cristina Gonzalez, Prince Gupta, Jiabo Hu, Yifei Huang, Yiming Huang, Weslie Khoo, Anush Kumar, Robert Kuo, Sach Lakhavani, Miao Liu, Mi Luo, Zhengyi Luo, Brighid Meredith, Austin Miller, Oluwatumininu Oguntola, Xiaqing Pan, Penny Peng, Shraman Pramanick, Merey Ramazanova, Fiona Ryan, Wei Shan, Kiran Somasundaram, Chenan Song, Audrey Southerland, Masatoshi Tateno, Huiyu Wang, Yuchen Wang, Takuma Yagi, Mingfei Yan, Xitong Yang, Zecheng Yu, Shengxin Cindy Zha, Chen Zhao, Ziwei Zhao, Zhifan Zhu, Jeff Zhuo, Pablo Arbelaez, Gedas Bertasius, David Crandall, Dima Damen, Jakob Engel, Giovanni Maria Farinella, Antonino Furnari, Bernard Ghanem, Judy Hoffman, C. V. Jawahar, Richard Newcombe, Hyun Soo Park, James M. Rehg, Yoichi Sato, Manolis Savva, Jianbo Shi, Mike Zheng Shou, and Michael Wray. Ego-exo4d: Understanding skilled human activity from first- and third-person perspectives, 2023.

Sheng Li and Handong Zhao. A survey on representation learning for user modeling. In Christian Bessiere (ed.), *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20*, pp. 4997–5003. International Joint Conferences on Artificial Intelligence Organization, 7 2020. doi: 10.24963/ijcai.2020/695. URL https://doi.org/10.24963/ijcai.2020/695. Survey track.

Xin Wang, Taein Kwon, Mahdi Rad, Bowen Pan, Ishani Chakraborty, Sean Andrist, Dan Bohus, Ashley Feniello, Bugra Tekin, Felipe Vieira Frujeri, Neel Joshi, and Marc Pollefeys. Holoassist: an egocentric human interaction dataset for interactive ai assistants in the real world, 2023.

Geoffrey I. Webb, Michael J. Pazzani, and Daniel Billsus. Machine learning for user modeling. *User Modeling and User-Adapted Interaction*, 11(1):19–29, Mar 2001. ISSN 1573-1391. doi: 10.1023/A:1011117102175. URL https://doi.org/10.1023/A:1011117102175.

Zihui Xue and Kristen Grauman. Learning fine-grained view-invariant representations from unpaired ego-exo videos via temporal alignment, 2023.

Mustafa Mert Çelikok, Pierre-Alexandre Murena, and Samuel Kaski. Modeling needs user modeling. *Frontiers in Artificial Intelligence*, 6, 2023. ISSN 2624-8212. doi: 10.3389/frai.2023.1097891. URL https://www.frontiersin.org/articles/10.3389/frai.2023.1097891.