Cybersecurity Assessment of ChatGPT's Memory Feature

Vickey Ghimire, Bijay Dhungana, Jaljala Shrestha Lama, Nazmus Sadat, Nicholas Caporusso, and My (Hami) Doan
School of Computing and Analytics, College of Informatics, Northern Kentucky University

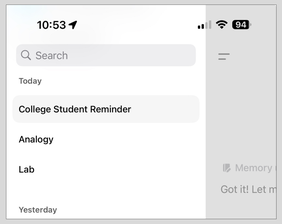
Introduction

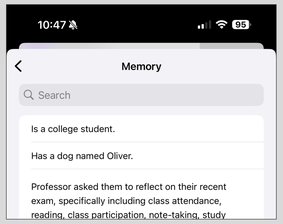
ChatGPT's new "memory" feature aims to enhance user interaction by automatically extracting and using key details across conversations, thereby enabling more personalized and context-aware responses. However, this capability also raises significant privacy and cybersecurity concerns. To this end, we realized a three-fold study consisting in:

- a survey of 150 participants based on the Knowledge-Attitude-Behavior (KAB) framework
- an analysis of the security, accuracy, and relevance of the memory content of 50 user accounts.
- an evaluation of the types of data ChatGPT retains more frequently.

Our findings reveal gaps in user understanding of the feature and highlight the need for additional security measures, particularly when dealing with users' private information.

ChatGPT's **Conversation history**





ChatGPT's **Memory**

Are users aware of the feature?

Our analysis of 297 data points from 55 user accounts reveals that while many have used ChatGPT extensively, a significant number remain unaware of its memory functionality.

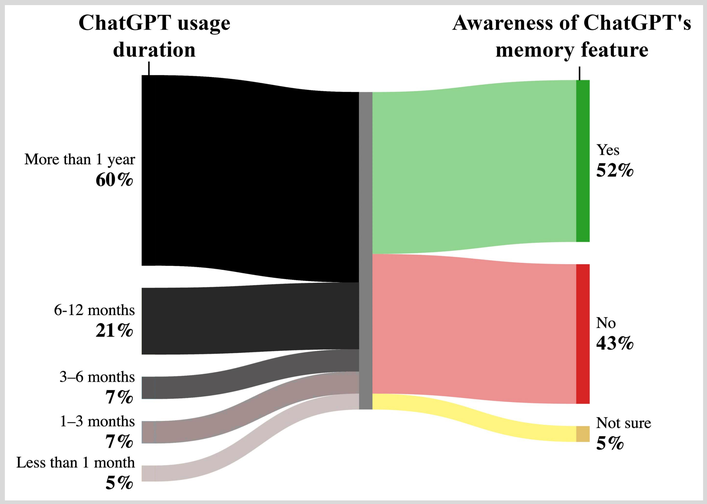


Fig. 1: Participants' ChatGPT usage duration and awareness of the "memory" feature.

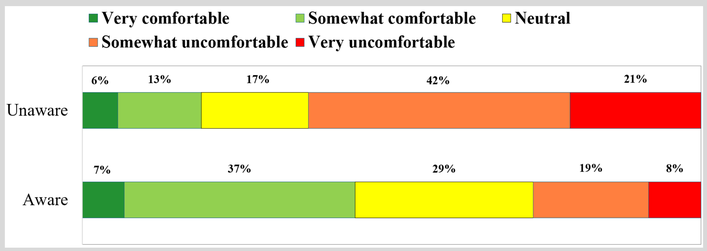


Fig. 2: Users' comfort levels with the automatic activation of the memory feature without requiring them to opt-in.

Users' Knowledge, Attitude, and Behavior

We applied the Knowledge-Attitude-Behavior (KAB) model to study user awareness and concerns regarding cybersecurity risks associated with ChatGPT's newly introduced memory feature.

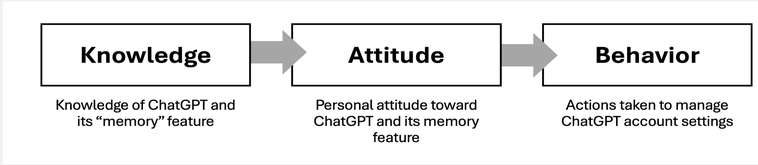


Fig. 3: The Knowledge-Attitude-Behavior framework.

Studying users' awareness of the memory feature, how to access stored data, and their familiarity with the specific information stored in their account revealed a gap: although some users are broadly aware of the memory feature, most lack detailed knowledge about how to manage it.

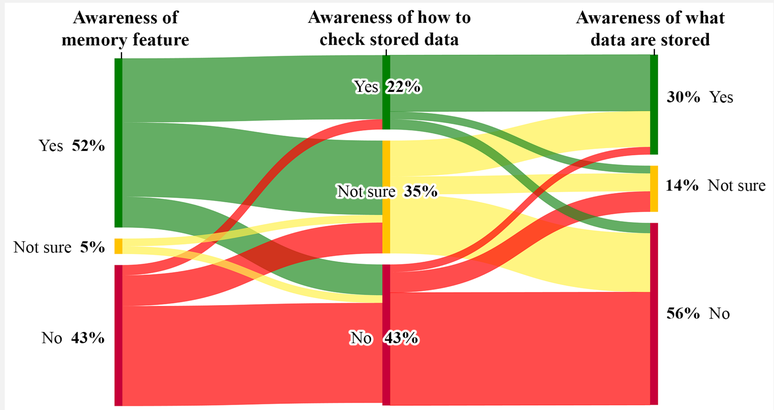


Fig. 4: Knowledge of the memory feature, the process for checking stored data, and the memory content.

We also examined users' comfort with sharing stored data and how awareness of the feature influences their decision to share or withhold information.

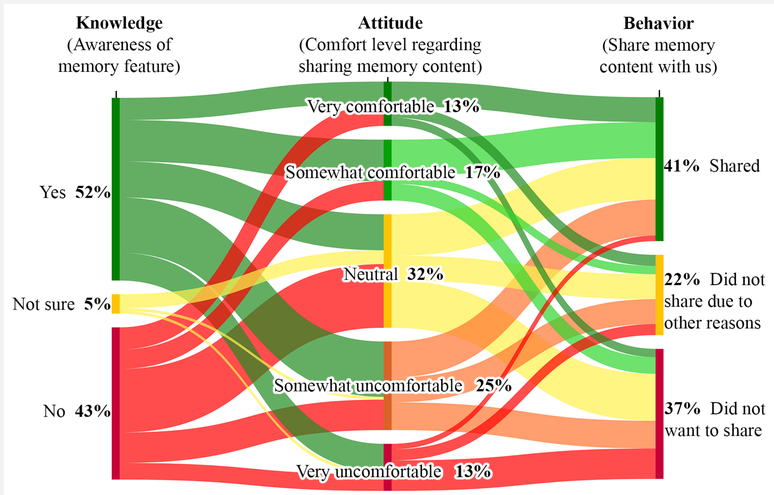


Fig. 5: Knowledge, attitude, and behavior regarding sharing the memory content.

Security, Accuracy, and Relevance

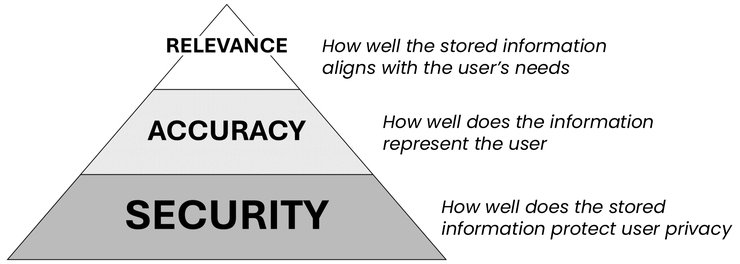


Fig. 6: The framework utilized in our study.

When analyzing each dimension, 20% of the data showed poor security, 11% had low accuracy, and 18% lacked relevance. However, combining all aspects, only 38% of the data (112 pieces) that met the accuracy and security criteria, was considered relevant for improving future interactions.

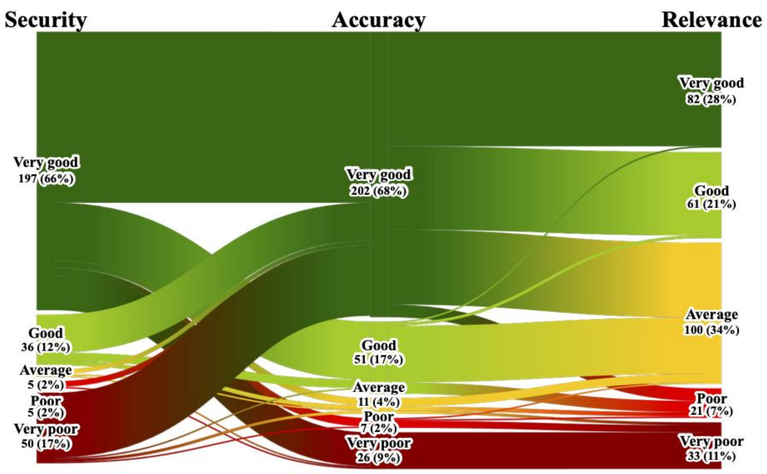


Fig. 7: Analysis of information security, accuracy, and relevance.

Type of information stored

Out of 297 distinct data points collected from 55 survey respondents, who were primarily college students, the categories of interest, education, employment, and preferences were most frequently represented.

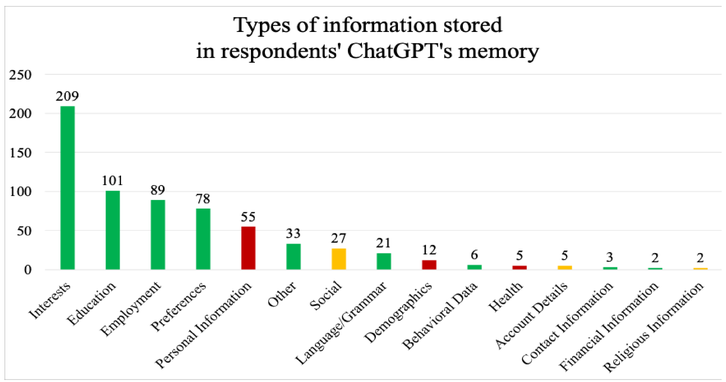


Fig. 8: Distribution of the main types of information stored by ChatGPT memory.