Working title: Data Descriptor for HMC Data Set

Angelica Henestrosa

Leibniz-Institut für Wissensmedien, Tübingen

October 14, 2025

Abstract

Since the emergence of large language models (LLMs) in 2022, generative AI has rapidly expanded into mainstream applications, leading, for example, to the integration of Apple Intelligence into customer devices in 2024. This integration into personal technology marks a significant shift and a further reduction in barriers to use, bringing advanced AI capabilities into everyday devices and making them accessible to private individuals. Thus, the use of generative AI–consciously or unconsciously–along with interaction through LLM-powered (voice) assistants and engagement with AI-generated content is expected to increase significantly. However, data that link this usage to psychological variables and track it over time remain scarce. This longitudinal study comprises the data from an American sample across six waves at two-month intervals between September 2024 and July 2025. It examines user behavior, attitudes, knowledge, and perceptions related to generative AI. Thus, this data set allows for future research on psychological and behavioral dynamics of AI use over time, offering insights into user engagement and the individual factors connected to it.

1 Background and Summary

The introduction of transformer architectures in 2017 marked a major breakthrough in natural language processing (NLP), enabling significant advances in machine learning (ML) and the development of large language models (LLMs). These models, trained on vast corpora of text data, have demonstrated unprecedented capabilities in generating coherent and contextually relevant language. A milestone in public engagement with generative AI (GenAI) was the release of ChatGPT in November 2022, which made LLMs widely accessible to non-expert users. Since then, millions of individuals have interacted with conversational agents and other GenAI tools, often regularly integrating them into everyday tasks such as writing, coding, learning, and decision-making (LIT). This widespread proliferation of AI technologies, coupled with their increasingly diverse applications and personalized user experiences, raises the questions on how psychological factors shape and might explain differences in AI adoption and usage. As AI systems become more adaptive and embedded in everyday life, understanding the determinants of usage intensity, behavioral patterns, and types of use becomes essential. Moreover, the field of AI is evolving at a fast pace, and user characteristics such as attitudes and trust are subject to change over time. Therefore, longitudinal research that captures temporal fluctuations in user traits and behaviors is crucial.

Therfore, this longitudinally designed data set aims to capture the evolving perceptions of opportunities and risks associated with AI, perceived capabilities of AI systems, attitudes toward AI, trust in AI, willingness to delegate tasks to AI, areas of application, (to be continued) and the interrelationships among these constructs over time and get some hints on causality. Longitudinal studies are more likely to find changes if there is a potential change trigger (Zhao et al., 2024)

Central questions are whether predictors of technology acceptance as well as technology use change over time, whether the perception of AI-Tools as tools vs. agents (if so: what type of role/relationship) changes over time, whether this perception is related to concepts like credibility, trustworthiness, or task delegation, and whether factors such as social presence of perceive anthropomorphism mediate such processes. We also explore the long-term effects of delegating tasks to AI Tools on outcomes like perceived self-efficacy (writing skills), loneliness, or cognitive self-esteem and explore the moderating role of personality.

This project is a joint project from the human-computer interaction group at the Leibniz-Institut für Wissensmedien in Tübingen (IWM). There are several (how many should we mention?) preregistrations from group members focusing on their individual subquestions. For an overview of the work packages and their research questions, please visit our repository [LINK]. Thus, this data descriptor may be used to examine research questions across the individual work packages, the possibility to extract and analyze specific subgroups or individual trajectories ignored in the work packages. Because the data set was collected shortly before the public release of Apple Intelligence on consumer devices, it offers a timely snapshot of user attitudes and behaviors at a pivotal moment in AI adoption. This context enhances the relevance of the data for understanding emerging patterns in human-AI interaction. Moreover, the findings may provide early indicators of how psychological variables such as trust, perceived usefulness, and willingness to delegate tasks relate to AI usage, potentially offering prognosis of similar developments in other countries.

2 Methods

2.1 e.g.: Participants and Data Collection

To examine those changes and relationships, an American sample mainly consisting of AI users (specify) was invited to participate in this survey at two-month intervals between September 2024 and July 2025.

This study targets an US-American sample due to Apple announcing to release its new AI platform Apple Intelligence in autumn 2024 (in the US due to the stricter regulations in the EU) and we expect many people to be exposed to this AI on their Apple devices. Data collection started at the end of August 2024?? (six waves, roughly one year).

* Prolific * Invitation * time and intervals * retention rate * second sample -> invitation of wave1 participants * focus on users -> exclusion of nousers without intention * ethics approval

2.2 e.g.: Measurements

* List of all measures by wave

We collected sociodemographic information, including, age, gender, educational level, and household income from all participants at wave 1.

3 Data Records

Data records for each of the six waves are available in csv format at (tbd) together with the R/python scripts for data anonymization, data cleaning, and data preprocessing. That is, firstly the data was anonymized by removing participants' Prolific IDs and unused variables, empty variables resulting from faulty questionnaire programming, and xy were removed. Thus (filename) represents the cleaned and anonymized raw data, including the single items of each measurement. Second, variable names were harmonized and scales were calculated, resulting an the preprocessed data set xy, ready for analyses across scales. Moreover, a codebook explaining variable abbreviations and containing information about the waves in which the variable was collected (what else?) is available at (tbd).

Table 1 provides an overwiev of the demographic variables over all six waves.

	Total N	User	Male	Female	Other	Age M(SD)	Education M(SD)	Income M(SD)
wave 1	1007	0.76%	500	494	13	38.68 (11.11)	4.37 (1.34)	3.55 (1.62)
wave 2	768	0.76%	375	384	8	39.37 (11.08)	4.33 (1.32)	3.55 (1.61)
wave 3	658	0.77%	318	332	6	39.86 (11.00)	4.30 (1.33)	3.57 (1.61)
wave 4	611	0.76%	282	323	5	40.13 (11.04)	4.22 (1.35)	3.50 (1.62)
wave 5	564	0.76%	259	300	4	40.43 (11.06)	4.19 (1.33)	3.48 (1.61)
wave 6	514	0.76%	238	270	5	40.36 (11.12)	4.15 (1.33)	3.43 (1.59)

Table 1: Demographic variables per wave

4 Technical Validation

Wave 1 was conducted shortly before iOs 18?? was published. -> were there any other external events potentially influencing the survey?

* Analysis of sample differences across waves -> was the sample equally distributed regarding sociodemographic characteristics?

* attention check * bot detection question * forced to respond

5 Usage Notes (optional)

Maybe here elaborate on limitations: * no data on no-users for wave 1-3 * not representative for age/gender/education/region due to focus on users * online survey: inattentive participants, fatigue effects especially in wave 1 and 6 (more variables) * rentention rate/dropout rate across waves

6 Code Availability

All python (version x) an R (version x) code for data anonymization, data cleaning, and preprocessing as well as the cleaned and the preprocessed data sets for each wave are stored in the public repository [link].

References

Author Contributions

Competing Interests

Acknowledgements

Hier ist ein R-Chunk:

```
> x <- rnorm(100)
> summary(x)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
-2.5079 -0.8108 -0.2314 -0.1425  0.6314  3.1375 
```