

RESEARCH STUDY PARTICIPANT INFORMATION SHEET

STUDY TITLE

“EVALUATING THE EFFECTIVENESS OF AN 'AWARENESS AGENT' IN
MANAGING DIGITAL INFORMATION OVERLOAD”

CONTACT DETAILS

Principle investigator: David Goddard ~ david.goddard@open.ac.uk ~ +44 7810 554581

Supervisor: Dr Paul Mulholland ~ paul.mulholland@open.ac.uk

INVITATION

You are being invited to take part in a research study. Before you decide whether or not to take part, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully.

We have asked you to participate in this study as someone with experience in IT or using information systems, who will be able to give informed input into the study process. While no specific knowledge is required on your part, we will ask you to adopt the role of a specified persona for the purpose of the study and to interact with simple IT systems.

Participation in the study is entirely voluntary and choosing to either take part or not take part in the study will have no impact on any relationship you may have with the institution of the Open University or the researchers. Study participants may also withdraw at any time without any negative consequences.

1. GENERAL INFORMATION

1.1 Purpose of the study

This study aims to explore and evaluate the effectiveness of an innovative digital tool, referred to as an 'awareness agent', in managing information overload in digital communication environments. The research is part of a larger effort to understand how advanced technologies, such as machine learning and large language models, can be harnessed to conduct research and improve user experience in managing digital information.

1.2 What does participation involve?

As a participant in this study, you will be engaging with a prototype of the awareness agent, which is designed to filter and present information based on user-driven preferences and criteria. Your involvement will include

interacting with the system, providing feedback on its functionality and usability, and participating in discussions that help us understand your experience with the tool.

1.3 Study phases

PHASE 1: INITIATION

You will meet with the researcher and together perform an initial setup of the study, after which you will independently conduct some actions to train the agent models that you have set up over the course of a couple of days.

PHASE 2: DATA GATHERING

Over the course of about a week you will be asked to use a web interface for a short time each day to review and classify incoming content processed by the agent. The purpose of this phase is twofold:

- To gather feedback from you about how well the agent works
- To continue fine-tuning the training of the agent over time

After phase 2 concludes, the researcher will spend a short time processing the data in readiness for phase 3.

PHASE 3: EVALUATION FEEDBACK

You will meet again with the researcher to start an additional evaluation. Since the completion of phase 2, the researchers will have used an AI large language model (LLM) to conduct an evaluation of the decisions made by the agent. You are asked to review the evaluations made by the LLM, so that we can understand its own performance in this task. The researcher will walk you through the web interface for this task and ask you to use it independently to review the LLM's decisions. You will not have to review every decision made by the LLM, but we ask that you cover at least a subset of them.

1.4 Duration of the study

Your participation in this study will span over a two to three week period. While continuous engagement is not required throughout this duration, there are several key stages where your involvement is essential. These include an initial setup session, daily interactions with the system for a short period, review meetings, and a final evaluation phase. The time commitment varies but is designed to be manageable alongside your regular activities.

1.5 Voluntary Participation

Your participation in this study is entirely voluntary. You have the right to withdraw at any time without any negative consequences.

2. WHAT WILL I BE ASKED TO DO IF I AGREE TO TAKE PART?

If you agree to participate in this study, we will ask you to participate in a number of activities over the course of approximately two-three weeks. A timeline of study activities is given below. While some activities are interactive with the researcher, others are intended to be carried out independently, and the overall study has been designed to fit alongside your usual activities with minimal disruption.

It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form.

3. TIMELINE AND KEY PHASES OF PARTICIPANT INVOLVEMENT

3.1 Day 1: Initiation session [phase 1]

- Duration: Approximately 1-2 hours.
- Activities: Orientation meeting with the researcher, agreeing data feeds and ML models for your persona, initial model training, Q&A session. Option to opt-out with no consequences.

3.2 Days 2-4: Early training and independent review [phase 1]

- Duration: Approximately 30 minutes daily.
- Activities: Independently review incoming content and adjust model classifications using the web interface. This phase is crucial for initial model training and adaptation.

3.3 Day 4: First review meeting [phase 1 completion]

- Duration: Approximately 1 hour.
- Activities: Follow-up discussion with the researcher, addressing any issues and completing a standardised questionnaire. Data from this point onwards is captured for the study.

3.4 Days 5-11: Continued daily usage [phase 2]

- Duration: Approximately 30 minutes daily.
- *It is not essential to participate every day, but any gaps should be kept to a minimum.*
- Activities: Continue reviewing and classifying incoming content independently.

3.5 Day 11: End of interaction phase [phase 2 completion]

- No active participation required.
- Researcher will conduct data analysis and prepare phase 2 content.

3.6 Day 14: Evaluation meeting [phase 3]

- Duration: Approximately 1-2 hours.
- Activities: Final meeting to review the study data and analysis, complete a questionnaire, and introduction to the evaluation explorer interface for independent review of the AI LLM's work.

3.7 Day 14+: Independent evaluation feedback [phase 3]

- Duration: Flexible, at participant's convenience.
- Activities: Use the evaluation explorer interface to provide feedback on the AI LLM's evaluation of the agent's classifications.

3.8 Completion of Participation

- Following the independent evaluation feedback, your active participation in the study will be complete. We appreciate your time and valuable contributions to this research.

4. HOW THE STUDY WORKS

4.1 Personas

PATHY TEMPLATES AND PERSONA ADOPTION

In this study, personas are constructed using PATHY templates, a structured framework that helps create detailed, realistic profiles representing potential users. These personas embody various characteristics, behaviours, and preferences that we use to inform our study of the agent's characteristics.

Participant's role: As a participant, you are expected to adopt a persona based on these templates. This involves understanding the persona's background, motivations, and information habits. Your ability to think and act in line with this persona is crucial, as it directly influences the study's outcomes.

TRANSLATING PERSONAS INTO DATA SOURCES

The persona you adopt will guide the selection of data sources. For example, a persona working in technology might require data feeds related to tech news, whereas a student persona might need academic resources. The aim is to mirror the information landscape the persona would realistically navigate.

Participant's role: You will play an active role in this process, working with the researcher to identify and agree on data sources that best represent your persona's information environment. Your insights and choices are valuable in customising the study experience to fit the persona.

4.2 Synthetic data

USAGE IN THE STUDY

Synthetic data is used to supplement real data sources, providing a controlled and ethically responsible way to simulate information scenarios relevant to each persona. This data is generated to reflect typical content a persona might encounter, such as social media posts, news articles, or work-related communications.

We aim to ensure that synthetic data is as realistic and relevant as possible, mirroring the complexity and nuance of actual data while avoiding privacy concerns associated with using real personal information.

ROLE OF THE PARTICIPANT IN DEFINING SYNTHETIC DATA

Collaborative creation: Your input helps define the themes, topics, and context for synthetic data generation. By understanding your persona, you will help guide the creation of data that is both meaningful and useful for the study.

Feedback and adjustments: You will have the opportunity to provide feedback on the synthetic data, ensuring it aligns with the persona's expected information environment. Your insights will be used to make any necessary adjustments to the data generation process.

4.3 AI LLM evaluation

PURPOSE AND METHODOLOGY

The study utilizes AI LLMs, like OpenAI, to evaluate the effectiveness of the agent's ML models in classifying information according to the persona's needs. This evaluation helps to understand the potential of AI in supporting information management. By using AI in this way we can potentially gain insight on a very large amount of data, but we also need to validate this technology itself so we can understand its limitations and most appropriate uses.

The AI will assess the classifications made by the ML models and provide its evaluation. We use both standard and fine-tuned versions of the AI models to compare their effectiveness in understanding and categorizing information accurately.

PARTICIPANT'S ROLE IN EVALUATION FEEDBACK

Feedback Mechanism: Towards the end of the study, you will be asked to review the evaluations that have been made by the AI evaluations. Using a specially designed web interface, you will provide feedback on whether you agree or disagree with the AI's judgments for a number of items.

Impact of Feedback: Your feedback is invaluable in assessing the alignment between the AI's evaluations and human judgment. It will help us understand the strengths and limitations of using AI for this purpose and guide future improvements in AI-driven studies and utilities.

5. PARTICIPANT INFORMATION PACK

As a participant you will be provided with the following information:

- Participant Information Sheet (this document)
- Your persona's PATHY template
- Participant interface user manual
- Technical details sheet
- Observations form

6. HOW WILL THE DATA I PROVIDE BE USED?

PERSONAL DATA

Your personal data, such as your name and contact details, will be gathered and used only for the purposes of administering the study and stored securely on Open University systems such as email servers. If you agree to receive follow-up communication relating to the study, then we will retain contact details for this purpose only; otherwise all personal data will be deleted on completion of your involvement in the study.

We will ask your consent to use a short, anonymous biography of you for use in documenting the research. We will only use text that has been approved by you for this purpose. An example is: "Participant 1 is an IT professional with 20 years experience in software development".

Due to the nature of the study, using synthetic and public data, we do not expect to gather any other personal data from you. However, we will conduct a post-study review to ensure that this is the case.

STUDY DATA

As a participant in the study, data will be gathered from you manually and automatically, as detailed below. The study itself will not gather any personal data however.

You will be asked to record observations about the study in text format. This information will be stored on a secure file server and retained for the duration of the research. Anonymised summarised/paraphrased extracts may be used in the final research report, after which the content will be destroyed.

Data is collected by computer systems during your interaction:

- Information about data feeds that you help define
- Values that you input for data items, where you select a value from an input box in the application user interface

This data is stored in a secure database in such a way that is not linked to you as an individual. This anonymous data will be retained for the duration of the research project and included in summarised form in

the final research report. Personal data such as IP address or your personal login name is not gathered or stored by any of the computer systems used in the study.

Some of the data used in the study is real data gathered from public external sources, such as newspaper feeds or social media. Some of these sources may refer to individuals, but will not be linked to you personally. All other data is entirely synthetic, created by an AI system according to rules defined for the persona that you are adopting. Public external and synthetic content gathered/generated during the study will be retained for the duration of the research. If there is any item of public or synthesised content that you do not want us to retain for any reason, we will remove it immediately on request.

YOUR RIGHT TO WITHDRAW FROM THE STUDY

You have the right to withdraw from the study at any time during your participation by contacting the researcher and/or ceasing to interact with the software.

You have the right to ask for your data to be removed after your participation in the study by contacting the lead researcher. This can be done until the data has been aggregated for analysis, which we anticipate to be by May 2024.

HOW DO I AGREE TO TAKE PART?

To opt in to this study, please return the signed consent form to the researcher, who will then schedule your involvement with you at a time of your choosing.

THANK YOU

Thank you for taking the time to read this information sheet. We appreciate your time and valuable contributions to this research.

DATA PROTECTION

The Data Controller for personal data provided by you is The Open University. The lawful reason for processing your data is that conducting academic research is part of The Open University's public task.

No data provided by you will be stored or processed outside of the UK/EEA.

While some data used in the research is processed by cloud service providers such as OpenAI, this does not include any personal data provided by you.

You have a number of rights as a data subject:

- To request a copy of the personal data we have about you
- To rectify any personal data which is inaccurate or incomplete
- To restrict the processing of your data
- To receive a copy of your data in an easily transferrable format (if relevant)
- To erase your data
- To object to us processing your data

If you are concerned about the way we have processed your personal information, you can contact the Information Commissioner's Office (ICO). Please visit the ICO's website for further details.