Study Protocol for Awareness Agent Testing

Using Software Agents to Raise Awareness and Lower Information Overload in a Multi-user Collaborative Environment

This document outlines the process for gathering experimental test results for the Awareness Agent. It covers general design of the experiment and the planned method of implementing it, as well as the expected results and how these will be used. The overall experimental approach is a technology/design probe, where individual test subjects are given supervised access to the test system, with case studies being generated from these interactions and subsequent interviews. Although qualitative in general approach, quantitative elements will be used to inform the conclusions.

Introduction and background

Introduction to information overload in digital communication

In an era increasingly dominated by digital communication, information overload has become a pervasive challenge. The deluge of data from social media, email, and various work-related instant messaging applications has led to a situation where managing the sheer volume and variety of information has become daunting for many users. This phenomenon, commonly referred to as 'information overload', is not just about the quantity of information but also concerns its quality, relevance, and the timeliness of its delivery.

Concept and purpose of the 'Awareness Agent'

Central to this research is the concept of an 'awareness agent', a novel approach designed to mitigate the effects of information overload. The notional awareness agent functions as a personal assistant, intelligently filtering and presenting information to users. It is envisioned as a tool that not only manages the flow of information based on user preferences and needs, but also empowers users by giving them control over the algorithms that determine what information is relevant to them. Unlike the algorithms used by social media platforms - which usually serve the interests of the platform first and foremost - the awareness agent prioritises the user's interests, potentially leading to a more balanced and less overwhelming information experience.

Aims and scope of the research

This research aims to find effective ways to balance awareness and information overload. The primary objective is to support users in processing the maximum amount of relevant information without succumbing to the negative effects of overload. This process, termed ‘information targeting’, involves refining and presenting information in a way that addresses the need to know while attempting to minimise overload. The study explores the development of a Java-based 'awareness agent' prototype, designed to demonstrate the concept and potential of such a tool in a real-world context.

The evolution of technology and LLM systems

The rapid evolution of technology, especially in the realm of Large Language Models (LLMs) such as OpenAI's ChatGPT, has significantly altered the landscape of digital communication. These advanced AI systems have introduced new dimensions to how information is processed, summarized, and presented, offering both challenges and opportunities in addressing information overload. The adaptation of these technologies in research and practical applications has opened new avenues for exploration, particularly in enhancing user experience in digital communication platforms. While the awareness agent as originally conceived was not dependent on LLMs but instead used a previous generation of machine learning (ML), we have sought to adapt and incorporate the new landscape in our research. We have done this in two ways: firstly by incorporating new LLMs directly into the application (to produce text summarisations of incoming content for example), and secondly by taking a novel approach of incorporating LLMs into the process of our research. We accomplish the latter in two ways: data synthesis, and evaluation of content and decisions.

Study Design

Methodological framework and study phases

This research adopts a mixed-methods approach, integrating qualitative and quantitative methodologies to provide a rounded evaluation of the awareness agent. The study is conducted in distinct yet interconnected phases, each employing specific research methods aligned with their objectives.

Initial phase: persona-driven environment development

In the initial phase, a participatory design approach is utilized to understand the personas and apply them to the agent. These personas, grounded in prior survey research, are critical in simulating diverse user scenarios. In collaboration with the researcher, participants engage in a process akin to role-playing, adopting a persona to guide the selection of relevant data sources and define the models and classifications used by the model. This phase leverages participatory techniques and our prior study to ensure that the personas accurately reflect plausible user behaviours and information interaction needs.

Data acquisition and model training: participatory design and iterative methodology

The subsequent phase of data acquisition and model training incorporates principles of participatory design, whereby participants actively engage in shaping the machine learning models. Through a combination of commercial and bespoke web interfaces, participants manually classify a selection of incoming items, thereby training the models. This iterative process of classification is itself an element of study in this process and provided qualitative data for analysis and comparison.

Evaluation using OpenAI: comparative quantitative analysis

A central component of the study is the comparative evaluation of machine learning model classifications by OpenAI. This phase employs a quantitative approach, where the performance of both the standard GPT model and a fine-tuned model is systematically compared with manual input from the user. The use of OpenAI for this evaluation introduces an element of experimental research, as it tests the hypotheses regarding the

efficacy of fine-tuning in ML models and the ability of LLMs to accurately assess classification decisions.

Final evaluation: mixed-methods analysis

The final evaluation phase uses a mixed-methods approach. We obtain quantitative data by performing a systematic comparison of the awareness agent's ML model's performance and OpenAI's evaluations, with human oversight and input providing a reference point. Concurrently, we gain qualitative insights by undertaking structured interviews with participants, focusing on their experiences and perceptions of the awareness agent's utility, interface design, and overall effectiveness. This phase integrates the quantitative findings with qualitative feedback, providing a holistic understanding of the research outcomes.

Participant Recruitment and Profile

Selection criteria and recruitment process

This study will engage a small, carefully selected group of participants. Recruitment will prioritize individuals who possess the ability to adopt a persona and engage with the 'design fiction' concept of the study. These participants will not only represent hypothetical users but will also contribute to the speculative envisioning of the awareness agent's application in various scenarios. Preference is given to technologically literate individuals with an interest in human-computer interaction (HCI) and innovative technological solutions, who themselves have experience of dealing with information overload.

Design fiction in research

The use of design fiction as a methodological tool is central to this study. Design fiction, as described in various academic works, is a heterogeneous set of methods and practices that enable the creation of diverse scholarly and design contributions [Evaluating Design Fiction: The Right Tool for the Job].

This approach allows for the speculative development of technology-based visions of future life and the creation of fictional worlds that encourage critical reflection among potential users [Design fiction is a critical design approach]. Design fictions have been used extensively to encourage reflection on technology matters, especially in HCI. This approach has been effective in fostering critical thinking and in defining requirements for novel devices and services [Design fictions for learning].

In this study, we will employ design fiction to envisage future technologies and their potential impact, thereby sparking critical reflection among participants.

Profile of Participants

Participants in this study will need to demonstrate an ability to think critically and creatively, especially in terms of understanding and adopting the personas. They should be able to see beyond the limitations of the current system and actively participate in the design fiction process. This requires a blend of imaginative thinking and practical understanding of technology. Participants do not necessarily need to be experts in technology, but a basic understanding of HCI concepts and a willingness to engage with the study's speculative approach are essential.

Data Source Definition and Synthesis

Interactive process of data source selection

The study begins with a collaborative and interactive process involving the study participant and the researcher to define the data sources relevant to each persona. This phase is crucial in ensuring that the data sources accurately reflect the hypothetical scenarios and information needs of the personas. Participants, guided by their understanding of the persona, work jointly with the researcher to identify and select public data sources that the persona would likely engage with. These sources typically include RSS feeds from news websites, social media platforms, and other public domains relevant to the persona's interests and daily life.

Defining synthesized data sources

In addition to public data sources, the study incorporates synthesized data, crafted to simulate realistic scenarios and interactions typical for the persona. This synthesized content is generated in batches by OpenAI, based on prompt text created by the participant in collaboration with the study administrator. The resulting content is stored in the AwAgData store and is served to the awareness agent application according to a defined schedule. With this approach we intend to deliver a rich and varied dataset that reflects the kind of information the persona would encounter in real life.

Machine Learning model definition and classification

Concurrent with the selection of data sources, the study involves defining specific machine learning (ML) models tailored to each persona. These models are designed to classify incoming information in ways that are meaningful and relevant to the persona. For instance, a model named 'personal-professional' might categorize items as 'personal' or 'professional', reflecting how the persona would differentiate between these spheres of life. The classifications made by these models are integral to the design fiction of the study. For example, an item classified as 'personal' by the 'personal-professional' model might be designed to be withheld from the user during work hours in a fully-realized application, unless it is also deemed 'urgent' by an 'urgency' model.

The role of design fiction in model interaction

The process of defining these models and their associated classifications is embedded in the study's design fiction approach. It requires participants to envision how a fully-realized awareness agent would interact with these classifications in the context of the persona's daily life. This approach allows for an exploration of the potential functionalities and user interactions of the awareness agent in a speculative yet grounded manner.

Data Collection and Analysis

Data collection infrastructure

Underlying the data collection process in this study is a bespoke back-end data service, referred to as "AwAgData". This service, utilising a REST-based interface, is used to record structured data throughout the study process. This data then stored in a backing SQL database, and other web services are defined that make it available in an appropriate form for study execution and analysis. The Awareness Agent application emits data to this service at critical junctures, such as data acquisition, classification, and evaluation events.

This approach ensures a comprehensive and accurate record of all activities related to the study.

Bespoke web interfaces for user interaction

The study employs three custom-designed web interfaces to facilitate various aspects of participant interaction and data management:

Training Interface

This interface allows participants to input data for training the machine learning models. Data arising from this interaction is passed to a back-end ML training/management API, and also to AwAgData for recording.

Evaluation Feedback Interface

Post-OpenAI evaluation, the feedback interface is used by participants to provide data about the evaluations conducted by OpenAI. Participants can agree or disagree with OpenAI's assessments, and their responses are recorded by AwAgData to be used in analysis.

Reporting and Analysis Interface

Used by the study administrator, this interface is used for the final reporting and analysis phase. It allows for the collation, review, and interpretation of data collected throughout the study.

Data capture through Slack integration

In addition to the data entered via the bespoke web interfaces, interactions within the Slack UI (such as user reclassifications of items) are also captured. The Slack Events API communicates these actions back to the Awareness Agent, which performs actions based on them (model training, moving items in the UI) and emits related data to AwAgData.

Analytical approach

We take both a quantitative and qualitative analytical approach. Quantitative analysis will focus on metrics derived from the data recorded by AwAgData. This includes the accuracy of classifications, the level of agreement between participant and OpenAI evaluations, and the efficacy of the machine learning models over time.

Qualitative analysis will draw from participant feedback obtained through structured interviews. This will provide insights into the user experience, the practicality of the awareness agent's interface, and the overall effectiveness of the system from the user’s perspective.

Ensuring data integrity and usability

Throughout the study, particular attention will be paid to ensuring the integrity and usability of the data collected. The structured and systematic approach to data capture, combined with the robustness of the back-end infrastructure, will ensure that the data is both reliable and amenable to in-depth analysis. All layers of the infrastructure have checks in place to identify and highlight data issues.

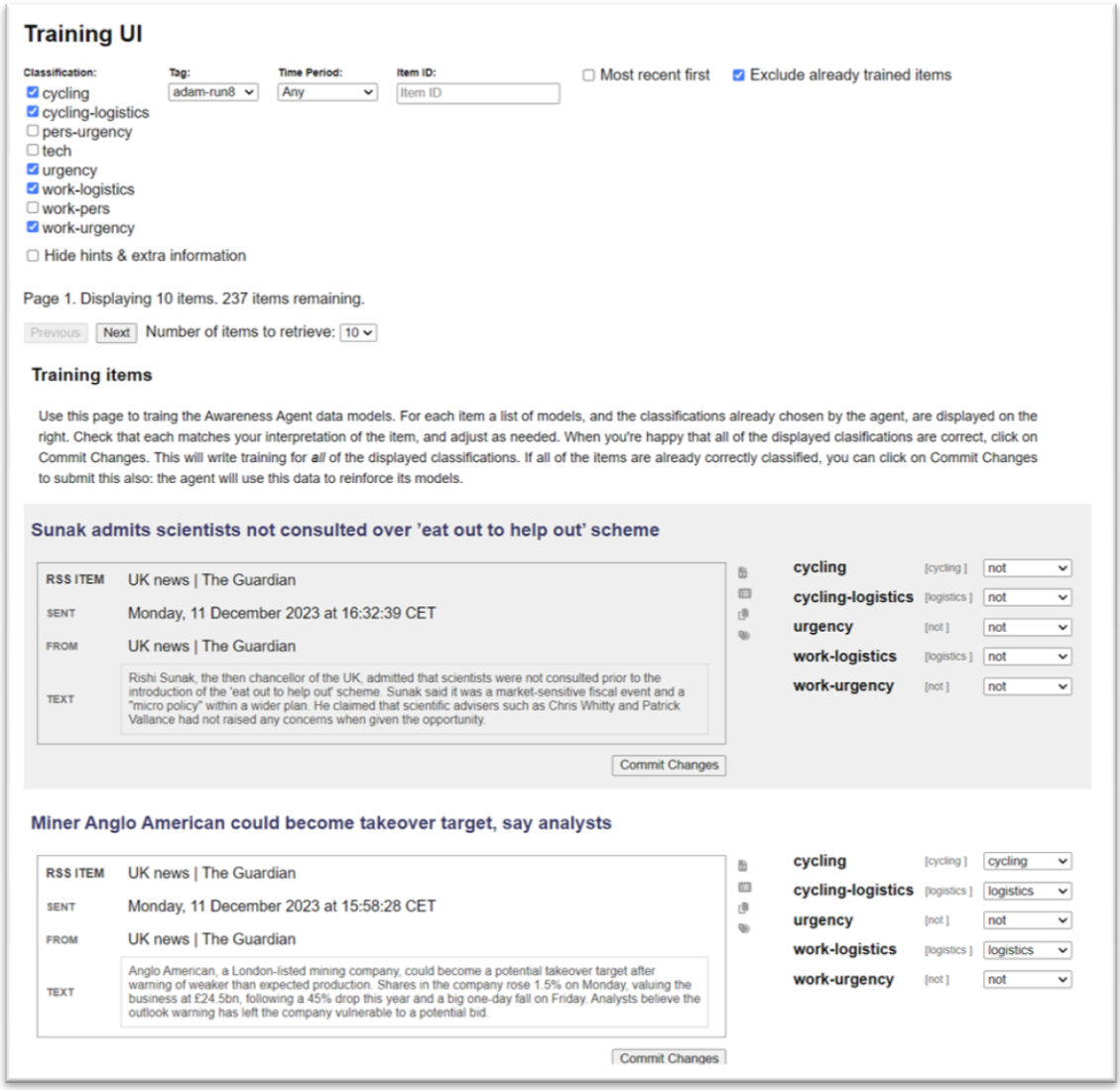

Figure 1 - Training/Classification UI

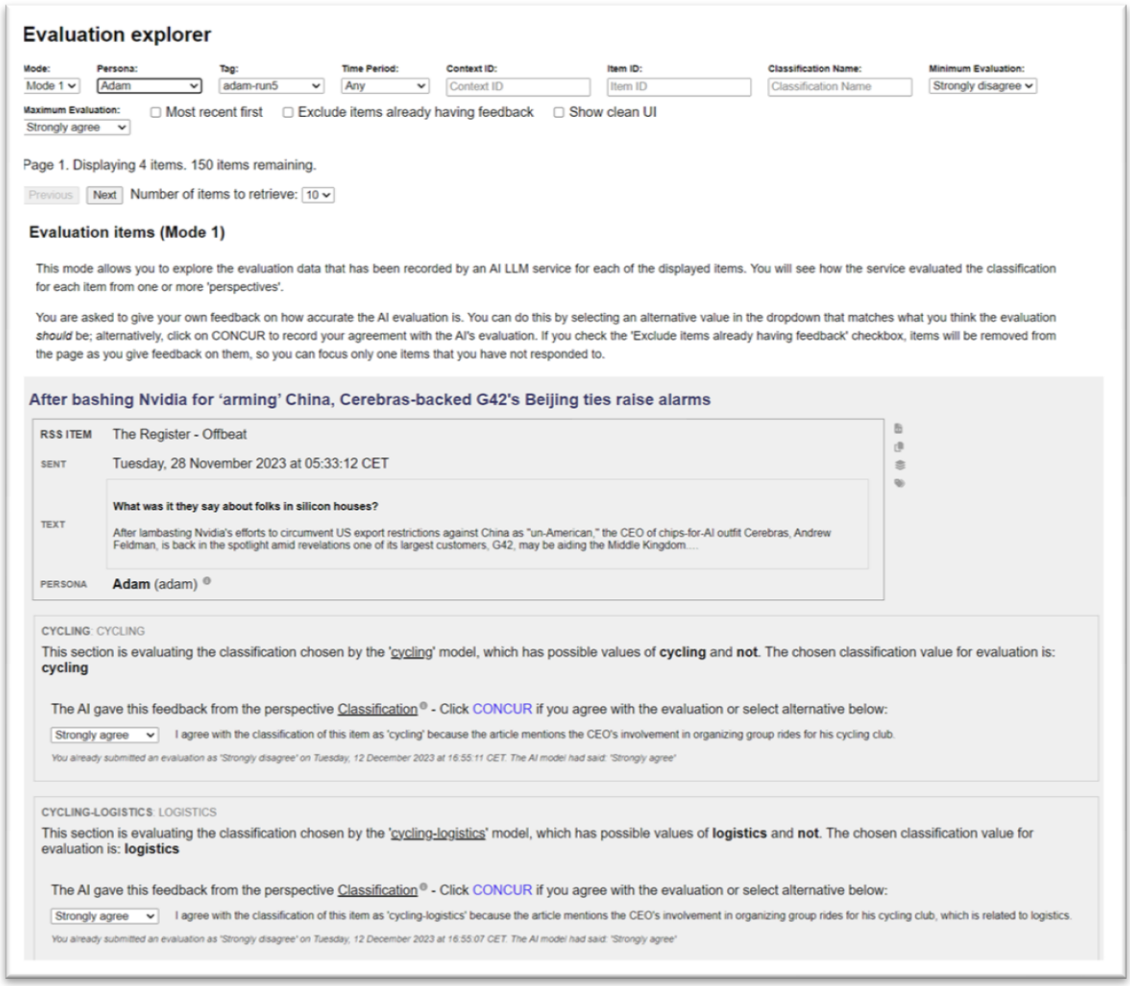

Figure 2 - Evaluation Feedback UI

Evaluation Metrics and Outcome Assessment

Quantitative metrics for evaluating the Awareness Agent

Accuracy of ML classifications (human assessed)

This metric assesses the accuracy with which the ML models classify the incoming information according to the persona-defined categories. It involves comparing the participant's manual classifications with the automated classifications of the models.

Accuracy of ML classifications (OpenAI assessed)

This metric also assesses the accuracy with which the ML models according to the persona-defined categories. In this case, it is generated by asking OpenAI to evaluate automated classifications of the agent's ML models as the adopted persona. This evaluation is performed using both vanilla and fine-tuned OpenAI models.

OpenAI evaluation concordance

This metric is generated from a human review of the OpenAI-assessed accuracy metric above. This metric will be crucial in assessing the effectiveness of OpenAI as a tool for validating and refining the ML models.

Improvement over time

Tracking the progression of the ML models' accuracy over the course of the study will provide insights into the learning capabilities of the models and the effectiveness of the training data.

Qualitative metrics for user experience and interaction

Participant Feedback on Interface Usability

Through interviews and the evaluation-feedback interface, participant responses will be gathered to assess the usability and user-friendliness of the awareness agent's interface as it exists, and as it is envisaged.

Perceived effectiveness of the Awareness Agent

Participants will provide insights into how effectively they believe the awareness agent aids in managing information overload, based on their experience interacting with the system.

User engagement with the Design Fiction

We will also evaluate participants' ability to engage with and contribute to the design fiction aspect of the study. This includes their effectiveness in adopting personas and envisioning how the awareness agent could function in a real-world context.

Outcome assessment

Comparative analysis of vanilla vs. fine-tuned OpenAI models

By comparing the performance of the standard GPT model with the fine-tuned model, the study will assess the impact of fine-tuning on the model's classification accuracy and relevance.

Effectiveness of synthetic data

Evaluating how well the synthetic data served the study's objectives will be essential. This includes assessing its realism, relevance to the persona, and its impact on the training and effectiveness of the ML models.

Overall system assessment

The final analysis will synthesize both quantitative and qualitative metrics to provide a comprehensive evaluation of the awareness agent's functionality, the efficacy of the ML models, and the utility of OpenAI's evaluations in a design fiction context.

Ethical Considerations

Informed consent and participant rights

The study will adhere strictly to ethical guidelines regarding informed consent. All participants will be fully informed about the study's objectives, methods, potential risks, and benefits before their participation. They will be required to sign a consent form acknowledging their understanding and voluntary agreement to participate. Participants will also be informed of their right to withdraw from the study at any point without any negative consequences.

Data privacy and confidentiality

Data privacy is a paramount concern in this study. To address this, the study primarily utilises synthetic data and public data sources, thereby mitigating risks associated with personal data privacy. Any data collected during the study will be anonymised and stored securely. Access to this data will be restricted to authorized personnel only, and it will be used solely for the purposes of this research in line with GDPR and other relevant legislation and good practice.

In cases where real data is used, strict measures will be implemented to ensure that the data does not contain personally identifiable information. The study's design inherently limits the exposure of sensitive or personal data by focusing on the persona rather than the actual participants.

Use of synthetic data

The use of synthetic data, generated through OpenAI, is a key ethical consideration. This approach not only serves the study's objectives but also provides an ethical advantage by reducing reliance on real personal data. The generation and use of synthetic data will be carefully monitored to ensure it aligns with ethical standards and does not inadvertently generate or replicate any unethical or harmful content.

Study sequence

Phase 1: preparatory and planning phase

Activities:

- Finalise study design and methodology.
- Prepare and test the technical infrastructure, including the AwAgData service and bespoke web interfaces.
- Develop and refine the selection criteria for participants.

Phase 2: participant onboarding and persona development

Activities:

- Recruit participants through targeted channels.
- Conduct initial meetings with participants to discuss the study and obtain informed consent.
- Collaboratively develop personas with participants and identify appropriate public and synthesized data sources.

Phase 3: data source definition and synthesis

Activities:

- Finalise the selection of public data sources and define synthesized data requirements.
- Generate synthesized data using OpenAI and store it in the AwAgData system.
- Define and set up the machine learning models for each persona.

Phase 4: data collection and model training

Activities:

- Begin data acquisition through the awareness agent.
- Participants engage in model training through manual classification of incoming items.
- Continuously monitor and record data interactions and classifications.

Phase 5: OpenAI evaluation and participant feedback

Activities:

- Conduct batch evaluations of ML model classifications using both vanilla and fine-tuned OpenAI models.
- Participants provide feedback on OpenAI's evaluations through the evaluation-feedback interface.

Phase 6: data analysis and reporting

Activities:

- Analyse the collected data, focusing on ML model accuracy, OpenAI evaluation concordance, and user experience.
- Conduct interviews with participants for qualitative feedback.
- Compile and write the final research report, summarising findings and insights.

Phase 7: review and write up

Activities:

- Review the study's findings with participants and stakeholders.
- Produce study write up for thesis chapter(s) and research paper.

Limitations and Challenges

Technical and resource limitations

Limited scope of Awareness Agent

Given the constraints of time and resources, the awareness agent developed for this study represents a limited prototype. Its functionality and the user interface do not fully capture the complexities and capabilities of a fully-realized application. This limitation might affect the depth of interaction and the range of insights obtainable from the study.

Restrictions of available APIs

The dependence on external APIs, particularly for data acquisition from social media and other platforms, poses a challenge. Limitations in these APIs may restrict the variety and volume of data that can be used, potentially impacting the representativeness of the data set.

Participant-related challenges

Engagement and persona adoption

The study's success heavily relies on participants' ability to effectively adopt and engage with their assigned personas. There is a risk that participants may not fully immerse themselves in the personas, which could influence the validity of the data collected and the study's overall findings.

Time commitment from participants

Given the interactive and iterative nature of the study, a time commitment is required from participants. This may limit the pool of potential participants and could affect the consistency of engagement throughout the study.

Data-related challenges

Synthetic data realism

While synthetic data is used to mitigate privacy concerns, there is a challenge in ensuring its realism and relevance to the personas. The quality of synthetic data directly impacts the training of ML models and the study's outcomes.

Balancing public and synthetic data

Finding the right balance between public and synthetic data is crucial. Too much reliance on either could skew the results and limit how well the study's findings can be generalised.

Methodological challenges

Interpretation of OpenAI evaluations

Relying on OpenAI for evaluating ML model classifications introduces a layer of complexity. The interpretation of these evaluations requires careful consideration, particularly in understanding how the OpenAI models' assessments align with human judgment.

Design Fiction approach

The use of design fiction as a research method is innovative but also untested in this context. There is a challenge in ensuring that this approach effectively simulates future scenarios and elicits meaningful participant interactions.

Expected Outcomes and Impact

Advancements in awareness-supporting technology

Insights into application design

The study is expected to yield significant insights into the design and functionality of an awareness agent. These findings will contribute to understanding the balance between user information needs and information overload, especially in a digital communication context.

Effectiveness of user-defined ML models

An important outcome will be an evaluation of how well user-defined ML models can manage and filter information based on individual preferences and needs. This could have implications for the development of more personalized and user-centric information management tools.

Contributions to AI and HCI research

Utilization of synthetic data and AI in research

The study will demonstrate the potential of using synthetic data and AI evaluations in research, providing a template for future studies in this area. It will contribute to the discussion on the ethical use of AI and synthetic data in research settings.

Novel use of design fiction in HCI

By employing design fiction as a research methodology, the study will explore new territory in HCI research. This could inspire further research that uses speculative and imaginative approaches to envision future technological applications.

Performance Assessment of OpenAI's LLMs

The study's findings on the performance of OpenAI's LLMs, both vanilla and fine-tuned, in evaluating ML classifications will be a key contribution. This will provide valuable insights into the capabilities and limitations of LLMs in tasks related to information classification and management.

Impact of Fine-Tuning on AI Performance

An analysis of how fine-tuning affects the performance of LLMs will offer important perspectives on the customization of AI tools for specific user needs and contexts.

Implications for future information management tools

Potential for real-world application

The study's findings could inform the development of future information management tools, especially those that seek to reduce information overload for users in both personal and professional settings.

Contribution to user-centric design principles

Insights gained from this study will contribute to the broader field of user-centric design, emphasizing the importance of user control and personalisation in technology design.