User Study Script

Welcome

Welcome to the user study. Let me start by introducing you to the background of this user study.

Background
In short, I am working on my Bachelors thesis titled "Interactive Debugging of LLM Agents for Software Development." As part of my thesis I have developed a tool to assist in the development of such LLM agents. The tool helps understand the behavior of an agent, and to identify possible bugs. By doing so, I hope that it can reduce the development time, thus increasing productivity. As the next step, I am conducting this user study to evaluate its usability and utility.

The Debugger UI
Here's a quick preview of the debugger.

Study Procedure
Our study does A/B testing, which is the simplest form of a between-group study. That means we have two conditions: one where the participant uses my tool, and one without. Each participant has been assigned to a group, and I will shortly reveal which group you have been assigned to.

We will start with a brief introduction to LLM agents, followed by a brief introduction to the tools available to complete the tasks later on. Depending on your group assigment, you will either use my debugger tool, or review raw log files in VS Code.

Following the introductions, you will be asked to complete task #1, in which you have to understand an agent trajectory and answer some questions about it. In task #2 and #3 you have to identify a likely bug in the agent program based on the trajectory. Each task will be timed, and they are devised so that you may run out of time before answering all of the questions. But, don't feel rushed, just solve as many as you can making sure to provide quality answers.

After each task, you will be asked to fill out the NASA Task Load Index (TLX) form, so we can get feedback on the subjective difficulty of each task.

Finally, if you were lucky enough to use our debugger tool, you will be asked to fill out a feedback form about your experience. This feedback will help me further improve the tool before its release.

Data Collected
For the purpose of the study, we will collect the following data about you:

- Your email (however, only used for correspondence and not published),
- Biographic information (needed to balance the two groups)
- Your answers as well as the time needed to complete the tasks
  Your data will be used for research purposes, and may be published in anonymized form identified by your participant key. This key is a random string and can not be used to determine your identity.

Consent

Before we get started, please sign our consent forms.

Start of Session

1. Background

2. Study procedure
3. Data collected
4. **Consent forms**

Introduction to LLM Agents

The goal is to use technology to solve complex conceptual tasks autonomously. A program that does that is referred to as an agent. In our case examples of such tasks may be to fix bugs in the code base, add new features, or develop and execute tests.

LLM agents work by employing Large Language Models (LLMs) for reasoning, decision making, content creation, etc. When you hear of an LLM just think about ChatGPT.

In summary, you can think of an agent as a program that is given some data with a task description, and uses ChatGPT along with some built-in tools to solve the task.

Agent Program
The tasks agents strive to solve are complex, meaning they can't be solved with one single prompt. For this reason, the solution is broken up into smaller steps. It is the agent program's job to facilitate a structured thought process and act as a middleware between the code and the runtime environment. Here's a graphical overview of how an agent operates.

Workflow of an LLM Agent
Agents work in cycles, also called steps. In each cycle, the agent figures out what to do next and informs the agent program. Next, the agent program executes the action, and appends the results to the history. Then the process begins from the beginning, and continues until either a solution is found, or the execution budget runs out.

We refer to the run of an agent as a trajectory.

Another Way of Looking at Things
Another way of looking at the workflow, is as a series of interactions between the LLM, agent program and the tools.

Introduction to the Tools

Next, we will review the tools at your disposal to solve the tasks. But first, I will reveal which group you have been assigned to.

- Reveal group assignment

Group A: Agent Debugger

1. Show general layout
2. Show agent activity chat
3. Explain bubbles
4. Show message inspector
5. Show resizing and movability
6. Show compare messages
7. Show repository panel
8. Show git commit tracking features
9. Show git commits in chat
10. **Provide 3 minutes for familiarization**

Group B: Raw Logs with Text Editor

1. Show raw linear log
2. Explain linear log and show steps
3. Show `.traj` JSON trajectory file.
4. Explain JSON and structure.
5. **Provide 3 minutes for familiarization**

Task #1: Agent Trajectory Comprehension

Your Task You will analyze a log of an LLM agent used for some aspect of software development. The log details the agent’s thoughts, actions, and results. You will answer a series of questions—some high-level (e.g., about the agent’s goal) and some detailed (e.g., about specific actions or errors)—to test your understanding of the agent’s behavior.

You will have 25 minutes to complete this task.

You will be given extra time fill out the NASA TLX questionnaire.
Warning: there are two pages of questions. The third page is the NASA TLX.
Warning: Mac keyboard shortcuts.
Warning: Keyboard Layout US or DE, please select.

Task #2 and #3: Agent Bug Localization Task

Your Task You will analyze the behavior of an LLM agent designed for software development to identify the likely causes of its failure to complete a given task. The goal is to pinpoint bugs in the agent's core code that led to the failure, rather than issues with the LLM's responses. Potential bugs may include misinterpretations of the LLM's outputs, incorrect logic in the agent's decision-making, or other implementation errors.

You will have 12 minutes to complete this task

You will be given extra time fill out the NASA TLX questionnaire.