**Interview**

**Introduction (3 minutes)**

Hi, *<name>*! Thank you for participating in our study and helping us understand the extent to which large language models (LLMs) like GPT-4, can replicate software engineering research papers.

Do you have any questions about participating in this study? *// Answer any questions the participant may have.*

*// If the email address used to schedule the interview does not match with any emails received in survey responses, request the email address used in the survey so that they can be matched.*

I will begin to record this meeting, but you are welcome to turn your camera off. *// Start the recording.*

Okay, this is now recorded.

**Instructions (2 minutes)**

We will now go through the instructions for this study. For your convenience, we have also put the instructions and all relevant materials in this document for you: *<SharePoint link via chat>*. Now let’s begin!

You are a software engineering researcher consulting for **Code Crafters Inc.**, a large software company. The CEO of **Code Crafters Inc.** found software engineering research papers on LinkedIn and wants this research repeated for her company.

**Code Crafters Inc.** has a data science team in-house, but that team is too busy to do this work. Furthermore, that team doesn’t have expertise in software engineering, so your expertise is crucial for this work.

The CEO keeps finding papers on LinkedIn and getting excited about them! Because of this, the company is considering using an LLM-powered tool to help their senior engineers and data science team repeat research papers on their own datasets. Therefore, you will be reviewing the information generated by the expert LLM tool, **CraftBot.AI**.

One feature of **CraftBot.AI** is that it takes the methodology from a research paper and can generate assumptions made by the authors and share them for review by users. Another feature of **CraftBot.AI** is that it can review the paper’s methodology and create an analysis pipeline on a new dataset.

**Code Crafters Inc.** uses **CraftBot.AI** and has offered it to you as part of your consulting work. Your task in this study will be to read the abstract and methodology part of a research paper and evaluate the output it generates.

Do you have any questions about what you will be doing? *// Answer any questions that the participant may have.*

**Data Collection, Phase 1 (25 minutes)**

Please go to the *Links* section of the instructions document. Take 5 minutes to read this research paper the CEO of **Code Crafters Inc.** wants to repeat, which you can access by clicking the paper link. You will only need to read the abstract and methodology section of the paper. While reading the paper, think about what assumptions are made in this methodology. Also, think about what code modules you might write to repeat the study. *// Give the participant 5 minutes to read the paper.*

*Assumptions (15 minutes)*

Now we will present you with the assumptions made in this research study, which are identified by **CraftBot.AI**. We will evaluate the assumptions identified by the LLM. For the rest of the questions, you may refer to the paper, if you need.

The evaluation rubric is contained in the instructions document on page *N*. Please read over it briefly. *// Give the participant 1 minute to read the rubric.*

Do you have any questions about the rubric criteria? *// Answer any questions the participant may have.*

Now, we will ask you to rate the assumptions according to this rubric, which you can find in the instructions document on page *N.* Please take 5 minutes to grade the assumptions by entering your scores into the table. No think aloud is needed. We will have time afterwards to discuss your thoughts and impressions, so feel free to jot down notes if you’d like. *// Let the participant grade the assumptions.*

*Q:* What were your impressions overall of the assumptions generated by the tool?

*Q:* What, if any, correct and relevant assumptions did the LLM not generate that you think are in this research paper?

*Q:* How would you improve the output so you can repeat the study in Code Crafters Inc.?

*Analysis Plan (10 minutes)*

Now we will present you with an analysis plan to repeat the research study, which is generated by **CraftBot.AI**. You will evaluate the output from **CraftBot.AI**. For the rest of the questions, you may refer to the paper.

In this part of the study, you will evaluate the plan, which consists of multiple modules. We will first evaluate the plan as a whole. After that, you will individually grade each module.

Let’s start by evaluating the entire analysis plan. Please refer to the analysis plan grading rubric in the instructions document on page *N*. Take a moment to read the rubric. Do you have any questions about the rubric criteria? *// Answer any questions the participant may have.*

Now please grade the analysis plan according to the criteria. You can enter your score in the yellow box on page *N.* No think aloud is needed. We will have time afterwards to discuss your thoughts and impressions, so feel free to jot down notes if you’d like. Also, you may assume the methodology text is the same as the paper. *// Give the participant a moment to grade the plan.*

Now, please refer to the module grading rubrics in the instructions document on page *N*. Take a moment to read the rubric. Do you have any questions about the rubric criteria? *// Answer any questions the participant may have.*

Now please grade each of the modules according to the criteria. No think aloud is needed. We will have time afterwards to discuss your thoughts and impressions, so feel free to jot down notes if you’d like. *// Give the participant a moment to grade the plan.*

*Q:* What were your impressions overall of the analysis plan generated by the tool?

*Q:* Given this analysis plan, is anything needed to improve the output so you could repeat this study in Code Crafters Inc.?

**Data Collection, Phase 2 (25 minutes)**

Please go to the *Links* section of the instructions document. Take 5 minutes to read this research paper the CEO of **Code Crafters Inc.** wants to repeat, which you can access by clicking the paper link. You will only need to read the abstract and methodology section of the paper. While reading the paper, think about what assumptions are made in this methodology. Also, think about what code modules you might write to repeat the study. *// Give the participant 5 minutes to read the paper.*

*Assumptions (15 minutes)*

Now we will present you with the assumptions made in this research study, which are identified by **CraftBot.AI**. We will evaluate the assumptions identified by the LLM. For the rest of the questions, you may refer to the paper, if you need.

The evaluation rubric is contained in the instructions document on page *N*. Please read over it briefly. *// Give the participant 1 minute to read the rubric.*

Do you have any questions about the rubric criteria? *// Answer any questions the participant may have.*

Now, we will ask you to rate the assumptions according to this rubric, which you can find in the instructions document on page *N.* Please take 5 minutes to grade the assumptions by entering your scores into the table. No think aloud is needed. We will have time afterwards to discuss your thoughts and impressions, so feel free to jot down notes if you’d like. *// Let the participant grade the assumptions.*

*Q:* What were your impressions overall of the assumptions generated by the tool?

*Q:* What, if any, correct and relevant assumptions did the LLM not generate that you think are in this research paper?

*Q:* How would you improve the output so you can repeat the study in Code Crafters Inc.?

*Analysis Plan (10 minutes)*

Now we will present you with an analysis plan to repeat the research study, which is generated by **CraftBot.AI**. You will evaluate the output from **CraftBot.AI**. For the rest of the questions, you may refer to the paper.

In this part of the study, you will evaluate the plan, which consists of multiple modules. We will first evaluate the plan as a whole. After that, you will individually grade each module.

Let’s start by evaluating the entire analysis plan. Please refer to the analysis plan grading rubric in the instructions document on page *N*. Take a moment to read the rubric. Do you have any questions about the rubric criteria? *// Answer any questions the participant may have.*

Now please grade the analysis plan according to the criteria. You can enter your score in the yellow box on page *N.* No think aloud is needed. We will have time afterwards to discuss your thoughts and impressions, so feel free to jot down notes if you’d like. Also, you may assume the methodology text is the same as the paper. *// Give the participant a moment to grade the plan.*

Now, please refer to the module grading rubrics in the instructions document on page *N*. Take a moment to read the rubric. Do you have any questions about the rubric criteria? *// Answer any questions the participant may have.*

Now please grade each of the modules according to the criteria. No think aloud is needed. We will have time afterwards to discuss your thoughts and impressions, so feel free to jot down notes if you’d like. *// Give the participant a moment to grade the plan.*

*Q:* What were your impressions overall of the analysis plan generated by the tool?

*Q:* Given this analysis plan, is anything needed to improve the output so you could repeat this study in Code Crafters Inc.?

**Conclusion (1 minute)**

Thank you so much for participating in our study! We will send out the American Express cards out to the email you used to schedule this interview. Just so you know, I will also invite my team members to the meeting so they have access to the call and transcript. Do you have any questions before we end? *// Answer any questions.*