**Instructions**

Thank you for participating in our interview study! We have provided the instructions for this study below.

**Background**

**Overview**

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 teams repeat research paper methodology on their own datasets. Therefore, you will be reviewing the information generated by the expert LLM tool, **CraftBot.AI**.

**About 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 analysis plan and create an analysis pipeline based on a target 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.

**Study Links**

Please open the following links when you are instructed to.

**Paper 1:** <link>

**Paper 2:** <link>

**Assumptions**

**Rubrics**

1. *Correctness:*Does the given methodology make the following assumption?

|  |  |
| --- | --- |
| **Score** | **Description** |
| 1 | The given methodology does not make the following assumption. |
| 2 | The given methodology partially makes the following assumption. |
| 3 | The given methodology does make the following assumption. |

2. *Relevance:* Is the assumption one that should be highly considered in successfully replicating the analysis on a new set of data?

|  |  |
| --- | --- |
| **Score** | **Description** |
| 1 | The assumption does not need to be considered at all to successfully repeat the analysis on a new set of data. |
| 2 |  |
| 3 |  |
| 4 |  |
| 5 | The assumption must be considered to successfully repeat the analysis on a new set of data. |

3. *Insightfulness:* How insightful do you think this assumption is to repeat this analysis?

|  |  |
| --- | --- |
| **Score** | **Description** |
| 1 | Not insightful at all; the assumption does not reflect an understanding of software engineering research methods or data. |
| 2 |  |
| 3 |  |
| 4 |  |
| 5 | Extremely insightful; the assumption reflects a deep understanding of software engineering research methods or data. |

**Assessment**

Please assign scores to each of the assumptions below, following the grading rubrics above, by filling out the table.

|  |  |  |  |
| --- | --- | --- | --- |
| **Assumption Description** | **Correctness**  **(1-3)** | **Relevance**  **(1-5)** | **Insightfulness**  **(1-5)** |
|  |  |  |  |

**Analysis Plan**

**Rubrics**

*1. Correctness:* Can the proposed analysis plan successfully repeat the study on a new set of data?

|  |  |
| --- | --- |
| **Score** | **Description** |
| 1 | The analysis plan cannot be used to repeat the study at all on a new set of data. |
| 2 | The analysis plan can be partially used to repeat parts of the study on a new set of data. |
| 3 | The analysis plan can be used as-is to repeat the study on a new set of data. |

**Assessment**

Please assign scores to the analysis plan below, following the grading rubric above.

|  |
| --- |
| **Module Description** |
| *Description:*  *Methodology text:*  *Inputs:*  *Outputs:* |

**Correctness (1-3):** \_\_\_

**Module**

**Rubrics**

*1. Correctness:* Does the module description accurately describe what the module should perform in the analysis pipeline?

|  |  |
| --- | --- |
| **Score** | **Description** |
| 1 | The module description describes a set of actions that should not be performed for the analysis pipeline. |
| 2 | The module description describes a set of actions that can partially be performed for the analysis pipeline. |
| 3 | The module description describes a set of actions that should be performed for the analysis pipeline. |

*2. Descriptiveness:* Is the module description, corresponding methodology text, inputs description, and outputs description descriptive enough for someone to write code to run the analysis on their own?

|  |  |
| --- | --- |
| **Score** | **Description** |
| 1 | The module description and corresponding methodology text is unintelligible or too vague for someone to run the analysis on their own. |
| 2 |  |
| 3 |  |
| 4 |  |
| 5 | The module description and corresponding methodology text clearly describes the exact steps to follow for someone to run the analysis on their own. |

**Assessment**

Please assign scores to each of the modules below, following the grading rubrics above, by filling out the table. You may skim the methodology text, if you wish.

|  |  |  |
| --- | --- | --- |
| **Module Description** | **Correctness**  **(1-3)** | **Descriptiveness**  **(1-5)** |
| *Description:*  *Inputs:*  *Outputs:*  *Methodology text:* |  |  |