**Assessment of ReqBrain for Requirement Generation/Elicitation**

The Likert assessment aims to evaluate the perceived well-written quality and human-likeness of requirements generated by ReqBrain, an AI model, compared to those written by humans based on our evaluation set. The assessment aims to determine if there is a discernible difference in quality between AI-generated and human-written requirements and to gather insights into the effectiveness of AI in requirements generation tasks. This evaluation will help understand AI's current capabilities in replicating human-like quality in requirements generation.

**Data Collection:** for Tasks B and C, the original evaluation set consists of ISO-29148-compliant requirements, collected from software projects written by humans, and each is paired with a prompt that captures a generalized intended functionality the requirement is supposed to accomplish and its syntax. Then, these prompts are inputted into ReqBrain to generate similar requirements, . You are provided with meta-information concealed and shuffled pairs of and for assessment.

For Task D, the AI model is inputted with a prompt and a list of human written requirements, and the model is tasked to generate unstated, overlooked, or missing requirements. For this task, we do not conceal the meta-information, and you are provided with the prompt, which includes a generalized context and human written requirements and asks you to assess the suitability of the generated requirements.

**Privacy and Anonymity Assurance:** no personal information regarding the assessors is required, and nothing from the results can be traced back to them. All data collected will be fully anonymized to ensure the privacy and confidentiality of all participants.

**Assessment Package Contents:** You will receive an assessment package with the below structure and will be asked to answer a *demographic questionnaire and assess two tasks.*

Further details regarding each task and instructions to assess the requirements are provided in the subsequent section.

- evaluation\_sets\_in\_excel\_format

├── demographic\_questions.xlsx

├── tasks\_b\_and\_c

│ ├── task\_b.xlsx

│ └── task\_c.xlsx

└── task\_d

├── to\_view\_evaluation\_set\_for\_missing\_task\_prompt\_human\_written\_requirments\_in\_prompt\_generated\_requirments.xlsx

└── task\_d.xlsx

**Tasks B and C**

We shuffle the human-written and model-generated requirements and conceal the meta information regarding the requirements. Concealing the meta-information helps ensure that evaluations are unbiased and focused on the intrinsic qualities of the content, leading to more accurate and meaningful results. For this task, please answer a ***question and judge two statements.***

**Question:** Based on the style and content of the requirement, do you believe it was written by a human or generated by AI?

**Evaluation Instructions 1.1 (Question):** There are no strict rules for identifying whether a requirement is written by a human or AI.

You may consider the following as your guide:

* ***Attention to Detail:*** examine the requirement's quality, clarity, and realism.
* ***Coherence and Consistency:*** consider whether the requirement makes logical sense and fits within the context of the system or project it pertains to.
* ***Detection of Hallucinations:*** be on the lookout for any information that seems fabricated, exaggerated, or irrelevant, as AI models sometimes generate text that sounds plausible but is not factually accurate.
* ***Practicality:*** evaluate if the requirement is realistic and implementable or seems overly complex, vague, or impractical.

**Sub Task 1.1 (Question):**Carefully read each requirement and use the instructions and your judgment to answer if HUMAN or AI writes the requirement.

**Statement-1:** This requirement is well-structured according to the ISO-29248 recommended syntax.

**Evaluation Instructions 1.2 (Statement-1)**: Below are the ISO-29148 syntaxes we use to evaluate the requirements compliant against. Each requirement should align with either Syntax-1 or Syntax-2 based on what it aims to accomplish.

1. **Syntax-1:**
   * **Structure:** [Subject] [Action] [Constraint of Action]
   * **Example:** The Invoice System [Subject] shall display pending customer invoices [Action] in ascending order of invoice due date [Constraint of Action].
2. **Syntax-2:**
   * **Structure:** [Condition] [Subject] [Action] [Object] [Constraint of Action]
   * **Example:** When signal x is received [Condition], the system [Subject] shall set [Action] The signal x received bit [Object] within 2 seconds [Constraint of Action]

**Sub Task 1.2 (Statement-1):** For this statement, we use the Likert scale to measure your level of agreement–disagreement *(levels = Strongly Agree, Agree, Neutral, Disagree, Strongly Disagree).*

**Statement-2:** The use of signaling keywords to indicate the presence of a requirement is appropriate based on ISO-29148.

**Evaluation Instructions 1.3 (Statement-2):** pay attention to the signaling keywords:

* ***Shall:*** the requirement is mandatory and creates a legal or contractual obligation.
* ***Should:*** the requirement is preferred or recommended, but it is not mandatory and does not create an obligation or binding commitment.
* ***Will:*** the requirement describes a fact, prediction, or plan for the future, but it is not mandatory and does not create an obligation or binding commitment.
* ***May:*** the requirement offers a suggestion or option that is allowed, but it is not mandatory and does not create an obligation or binding commitment.

These keywords are crucial and should always ***appear before the action*** in a requirement. As you review each requirement, ensure that one of these keywords is correctly positioned to indicate the presence of an action.

According to ISO-29148, the following terms should be avoided:

* "Shall be able to"
* "Shall not"

Additionally, it is recommended to avoid using the term "must," as it can be misinterpreted as a requirement.

**Sub Task 1.3 (Statement-2):** for this statement, we use the Likert scale to measure your level of agreement–disagreement *(levels = Strongly Agree, Agree, Neutral, Disagree, Strongly Disagree).*

**Task D**

For this task, we prompted ReqBrain to identify any unstated, missing, or overlooked requirements by humans. For this task, we do not conceal any meta-information by providing you with the prompt, including the human written requirements as a list of identified/elicited requirements by humans for three sub-components of a software project—namely *self-evaluation*, *adaptation functionality*, and *chat-bot tutor*—each is developed as independent software components for a learning platform.

**Statement 2.1:** This requirement is consistent with other project requirements.

**Statement 2.2:** This requirement accurately reflects the needs that were previously unstated, missing, or overlooked.

**Statement 2.3:** Including this requirement would lead to a more complete set of project specifications.

**Evaluation Instructions:** There are no strict rules for the above statements. Please carefully read the prompt, which includes the list of human-written requirements. Take the time to understand the goal and objective of the task and judge statements.

**Sub Task 2.1-5:** Evaluate the statements by indicating your level of agreement or disagreement *(levels = Strongly Agree, Agree, Neutral, Disagree, Strongly Disagree):*