Interview Protocol

[Introduction] In this study, we explored the capabilities of LLMs in generating Design Rationale (DR) for architecture design decisions using three different prompting strategies (zero-shot, CoT, and LLM-based agents). Although we conducted a thorough evaluation of the quality of the generated DR, it remains unclear whether such DR generated by LLMs is trustworthy and applicable in practical architecture decision-making. Therefore, we plan to interview practitioners to answer five interview questions (IQs) based on the 10 selected architecture problems that were provided to them prior to the interview.

[Interview Questions] The five interview questions of this interview are based on RQ4 in our study, that is, *How do practitioners perceive LLM-generated DR of architecture decisions in practice?*
Rationale: If the performance of LLMs in generating DR for software architecture is acceptable, their use for automatic DR generation could provide practical value for software architects. However, the perception of LLM-generated DR from the practitioners' perspective remains unclear. To address RQ4, we conducted semi-structured interviews to investigate how practitioners evaluate the DR generated by LLMs in terms of trustworthiness and applicability.

[Participants] The participants in this interview are practitioners with several years of experience in software development and architecture-related work. Their backgrounds ensured both practical knowledge of software engineering and familiarity with architectural decision-making processes.

[Data Analysis] For the answer of interview questions from practitioners, we will use open coding and Constant Comparison method to analyze and categorize the qualitative response.

Interview Steps:

Step 1: Before the interviews, participants are introduced to the research objectives and methodology of our study. They are then asked to review five randomly selected architecture problems and decisions, along with the DR provided by human experts and those generated by LLMs.

Step2: The interviews will be conducted by the first author. Each participant need to provide a brief overview of their background, including their professional role, as well as their years of experience in software development and architecture design.

Step3: The first author then conduct semi-structured interviews comprising five open-ended IQs. With participants' consent, the interviews are audio-recorded to facilitate subsequent analysis.

Interview Questions:

1. Trustworthiness

IQ 1.1: To what extent would you trust the DR generated by LLMs for architecture decisions? Why? You may choose from the following levels of trust:

- Do not trust it at all.
- Trust it only with substantial human review
- Trust it with some caution
- Generally trust it
- Fully trust it without human validation

IQ 1.2: What types of information (e.g., project context, domain-specific knowledge) should be provided to LLMs to enhance the trustworthiness of the generated DR for architecture decisions? Why?

2. Applicability

IQ 2.1: Would you consider using LLM-generated DR in practice? You may choose from the following levels:

- Definitely not
- Probably not
- Not sure
- Probably yes
- Definitely yes

IQ 2.2: What scenarios (e.g., comparing architectural alternative solutions, understanding architecture design) might be suitable for applying LLM-generated DR for architecture decisions?

IQ 2.3: What are the main challenges or limitations in applying LLM-generated DR in practice?