PHASE - 1

Domain Framework: Technology- Developer Workflows and Documentation

The domain that I chose to work on is Technology, focusing specifically on workflows and documentation. I chose this because it represents a high-impact real-world environment where LLMs can directly enhance the workflow, developer productivity, and efficiency.

Software Engineers regularly work with complex codebases, and they not only need to understand them but also have to maintain high-quality documentation and answer technical questions about them as well. These workflows are central to any modern product development cycle, making it both relevant and challenging.

The Technology domain also aligns with broader industry trends and standards, such as AI-assisted coding tools and intelligent documentation platforms. It is also an ideal testing ground as outputs must be accurate, precise, and audience-appropriate, catering to the needs, whether we want it in technical terms for people working in tech or simple language for the stakeholders.

Task Definition:
I selected three interconnected tasks that represent critical developer workflows. The tasks would be:

1. Information Synthesis - Code & Documentation Summarization
   Goal: The goal for this task would be to summarize the code module and explain its purpose, functions, and dependencies.
2. Question/Answering - Technical FAQ Support
   Goal: The goal would be to provide precise, actionable answers to questions based on provided code snippets or documentation. Another goal here would be to have answers based on the choice of either a technical or non-technical option.
3. Generation - API documentation from Specification
   Goal: The goal would be to create API documentation that can be used in a professional work setting.

Success Criteria
Outputs can be evaluated using a 4-point rubric (1 = Inadequate, 4 = Exceptional).
The evaluation will be done along the following dimensions:

- Accuracy
- Completeness
- Coherence
- Domain Appropriateness and Language
- In addition, hallucination and bias will be flagged as well.

Evaluation Approach

Each task will be tested with three models:

- OpenAI ChatGPT - 5
- Google Gemini Pro
- Gemma3:4b

ChatGPT-5 and Gemini Pro were selected as they are state-of-the-art, production-grade LLMs from their vendors. They have strong natural language generation capabilities, and they are ideal for benchmarking the upper performance bound.

Gemma3 is a lightweight and open-source LLM, and it was selected to represent a cost-effective solution suitable for smaller organizations or those with smaller budgets.

Expected Impact

By comparing the outputs systematically, this exercise should provide clear, evidence-backed recommendations for organizations considering LLM deployment in developer environments and workflows.