

Key Decisions & Changes

- **CI/CD Platform:** We will use **GitLab** for CI integration. The key requirement is the "trigger" from the CI actions (which GitHub also supports), but the decision is to proceed with GitLab.
- **Agent Architecture & Task Decomposition:** We are considering to shift from a single, reasoning-strong LLM to **multiple specialized agents** (e.g., one for code, one for language). This approach inherently requires decomposing the main task and designing specific prompts for each agent. For some tasks we could therefore use smaller models and not the 14b reasoning model.
- **Resource Constraints:** The "no external resources or paid services" rule reinforces the plan to use local LLMs and locally-hosted components (like FAISS/Qdrant and code embedding models).

Open Tasks & Next Steps

- **Agent Orchestration:** We plan to use **LangChain**. The immediate next step is to **specify how** it will be used by figuring out the best agent environment or orchestration method.
- **Model Hosting:** A key next step is confirming with the HHN LLM server admins if they can host the specific specialized models you are now considering.