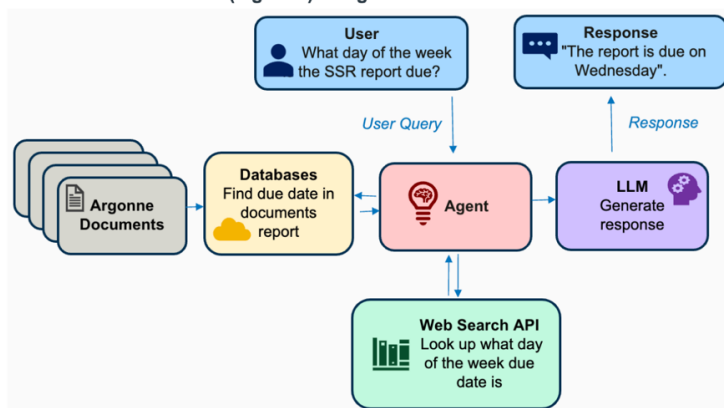


Argonne National Laboratory is developing an LLM-based chatbot named Argo for internal use. The chatbot's goal is to answer Argonne-specific questions based on information retrieved from proprietary documents. The team integrated an AI agent into Argo's existing capabilities and developed algorithms that can be used to benchmark its performance across multiple domains and quantitatively measure the impact of future improvements.

The team developed an AI agent allows Argo to autonomously combine externally sourced information internal documents as shown in the diagram (Figure 1). The team implemented a query routing system that routes domain-specific queries to fine-tuned LLMs that enhance Argo's ability to respond to queries related to math and coding. They also developed benchmarking tools that evaluate the effectiveness of Argo's new agent and various Large Language Model (LLM) backends. The agent's performance was scored across specific knowledge and reasoning categories to establish consistent and representative benchmarks that can be used to compare the existing architecture to future improvements. (Figure 2).

The team delivered an AI agent that is compatible with locally hosted LLMs, scripts to generate benchmarking reports on Argo, and a query routing system. Future steps involve integrating additional functionalities to the AI agent Argo, prompt engineering, and expanding the agent's capabilities. These advancements aim to enhance Argo's ability to provide accurate, reliable, and informative responses to users.

(Figure 1) AI Agent Architecture



(Figure 2) Agent Benchmarking Results Across Categories

