LLM4Ops Benchmark System Implementation

Presented by Junjielong Xu

Abstract –Recent researchers have attempted to explore the use of Large Language Models (LLMs) for IT Operations (LLM4Ops), including LLM-based microservice anomaly detection, fault diagnosis, and root cause analysis. While existing methods have shown decent performance in various enterprise services, the close-sourced nature of these services makes it difficult to compare the performance of different LLM4Ops methods. One potential solution is to collect offline data from enterprises and de-identify it. However, the offline nature of the data limits LLM4Ops research to only the anomaly detection or fault diagnosis, without delving into the resolution of these issues. To overcome this limitation, we have modified and implemented a testbed system with 24 microservices based on a train-ticket system in our lab's servers via Kubernetes. This system provided the LLM agent with full control over the system's APIs and commands, allowing researchers to evaluate different LLM4Ops methods in an online environment.