

BI4LLMC GROUP 2

LLM Interpretability & Explainability

1. Do current benchmarks adequately evaluate the interpretability and explainability of LLMs in coding tasks?

Interpretability in Agentic AI

1. Does the interaction between humans and multi-agent LLM-based SE systems necessitate a new theory of interpretability?
2. To what extent can interpretability techniques from LLM4SE be applied to single-agent systems in software engineering?
3. To what extent do single-agent interpretability techniques work in multi-agent SE?
4. How does interpretability change in multi-agent SE systems where agent decision-making is distributed?

Evolution of Agentic Benchmarks

1. What processes should govern the evolution and updating of agentic software engineering benchmarks to keep pace with rapid technological advancement while maintaining consistency and comparability across studies?