

Evaluation Performance: Rule-Based vs LLM Judge

