Evaluation Performance: Rule-Based vs LLM Judge

103 - Rule-based (This Work)

