

Average LLM-based Evaluation Scores per Dimension for Four LLMs Using DeepSeek-V3 as Evaluator

