

总览: [\(99+ 封私信 / 81 条消息\) 全景解读 LLM Agent 评测方法 - 知乎](#)

调用大模型去评估: [ToolBench/README_ZH.md at master · OpenBMB/ToolBench](#)

大部分采用的指标: [评测指标 | FlagEval](#)