对话模型 更新时间: 2025-03-26

Qwen2.5-Max是阿里云推出的旗舰语言模型,基于20万亿Token预训练数据,在数学推理、编程等任务中性能优异,支持长文本生成与多语言处理,并通过TensorRT-LLM优化提升部署效率。

论文 发布者 官网 参数 权重 阿里巴巴 qwen2.5 max qwen2.5 max 32B qwen2.5 max 性能维度 语言维度 综合 安全性 推理速度 计算资源消耗 中文 英文 推理精度

安全领域能力

