

LLMs 对比篇

来自：AiGC面试宝典

宁静致远

2024年01月27日 19:44



扫码
查看更

LLMs 训练数据 和 数据量 对比如何？

模型	训练数据	数据量
LLaMA	以英语为主的拉丁语系	1T/1.4T tokens
LLaMA2	以英语为主的拉丁语系	2T tokens
Falcon	英语、法语	1.5T tokens
Bloom	46种自然语言和13种编程语言	350B tokens
ChatGLM-6B	中英双语，中英文比例为1:1	1T tokens
ChatGLM2-6B	中英双语，中英文比例为1:1	1.4T tokens
baichuan	中英双语	1.2T tokens