????LLM???

??????????????

???????

现有模型的推理能力已经足够强，真正缺的是稳定的长期记忆。

我们需要那么多智能体协作，恰恰是因为一个智能体记不住所有事情。

????

- AI 缺的未必是更多上下文，很多时候缺的是上下文之间的关系。
- top-K chunks 可以把文档捞出来，但解释不了它们为什么重要、彼此为什么相连。
- 向量搜索解决的是 “像不像”，图更擅长回答 “为什么是它”。
- GraphRAG 不是替代向量检索，而是让检索命中之后还能顺着关系继续往下走。
- 当 AI 开始进入权限、历史、作者、决策轨迹这些复杂地带，纯文本上下文很快就会不够用。
- 图数据库重新变重要，不是数据库品类回潮，而是 AI 系统开始需要一层更像知识结构的上下文底座。

LLM????????

1. LLM大模型的权重
    1. 能进行一定深度和广度的处理能力，但是整体深度还是达不到人脑
    2. 需要通过及其复杂的梯度下降，不能快速存储特定数据，更新困难、无法追溯
    3. 黑盒，不确定性高，“增删改”不方便
    4. 计算资源消耗大，虽然有 MoE
2. KVCache 等运行态中间变量
    1. 存储密度比较低，存储容量上限低
        1. 相对于权重表示，KVCache依赖自然语言的 token，表达能力（复杂抽象不能用语言表达）和效率都差一点
    2. 逻辑推理能力稍差，思维能力稍差

1. CoT可能出现前后矛盾的表达，长上下文之后，模型的准确性下降
3. 计算资源消耗稍大
1. 潜式思维链（ Latent Chain-of-Thought）是一个改进方向
3. 外部专用数据存储系统
1. 容量大
2. 抽象层级最低 ，存储密度最低，能力差
3. 资源消耗小
4. 支持动态 “增删改 ”

Revision #2
Created 2026-05-05 14:38:32 UTC by Colin Updated 2026-05-16 04:11:08 UTC by Colin