

Perguntas de Reflexão

Como você equilibraria o uso de LLMs (que podem ser custosos) com métodos tradicionais de recomendação?

Uso LLMs quando o benefício justifica o custo. Para tarefas bem resolvidas com algoritmos tradicionais, mantenho o que é simples e barato. Reservo o LLM para onde ele realmente agrega, como personalização textual, contexto e geração natural de linguagem. Cache é essencial nesse equilíbrio.

Quais são os principais desafios éticos de usar IA generativa em um sistema de recomendação?

O principal risco é a geração de conteúdo enviesado ou inadequado. Modelos como esses não têm noção de contexto real, então podem reproduzir padrões que a gente não quer. Por isso, é importante prever fallback, validações e, se possível, revisão.

Como você abordaria o problema de explicabilidade em um sistema que combina métodos estatísticos com outputs de LLMs?

Separando bem as camadas. A recomendação é explicável, baseada em dados objetivos. A parte do LLM é mais aberta, então encaro como camada complementar — uma forma de enriquecer a experiência, não de justificar a recomendação.

Quais métricas você usaria para avaliar o sucesso de sua implementação em um ambiente de produção?

Como engenheiro, focaria em métricas técnicas: latência, disponibilidade, acertos de cache, quantidade de fallbacks. Mas pensando em valor gerado, a taxa de conversão de produtos recomendados em compras me parece o principal indicador.