

Una pipeline de LLM organiza las etapas del procesamiento de texto usando un modelo de lenguaje. Incluye entrada del usuario, preprocesamiento, recuperación de contexto, consulta al modelo y postprocesamiento de la respuesta. Esta arquitectura modular permite integrar componentes como bases vectoriales, APIs externas o verificadores de hechos. Además, facilita el monitoreo, la auditoría de salidas y el control de calidad. También permite experimentar con diferentes configuraciones de prompt...