

heise+ | KI-Server hosten: Mit vLLM KI-Modelle produktiv betreiben

2025-07-25 11:30

Lokale KI bietet Vorteile wie Kostenkontrolle und Datensicherheit. Für die Entwicklung eignet sich Ollama, doch für den Servereinsatz ist vLLM die bessere Wahl.