CZY LOKALNY LLM SPRAWDZI SIĘ W MOJEJ ORGANIZACJI?

Korzyści korzystania z lokalnych modeli LLM

Obszar	Osoba indywidualna	NGO
Prywatność	Kontrola nad danymi (zostają na komputerze)	Kontrola nad danymi. Instalacja na własnym serwerze, VPS-ie lub serwerze dedykowanym, bez przesyłania danych do zagranicznych centrów.
Koszty	Brak opłat za użycie	Stałe, przewidywalne koszty
Dostęp	Działa offline, możliwość dostępu zdalnego z własnym serwerem	Działa offline, możliwość dostępu zdalnego z własnym serwerem
Możliwości	Możliwość testowania wielu modeli.	Możliwość dopasowania do procesów i potrzeb organizacji (np. system RAG na własnych dokumentach), fine-tunnig (dopasowanie modelu) do konkretnych branż czy zadań. Możliwość testowania wielu modeli. Możliwość wykorzystania małych modeli (Small Language Models) do konkretnych zadań.
Niezależność	Brak zależności od dostawców. Brak ryzyka zmian cen/polityk platform.	Brak zależności od dostawców. Brak ryzyka zmian cen/polityk platform.
Rozwój kompetencji	Lepsze zrozumienie działania AI	Budowanie kompetencji zespołowych, tańsze eksperymentowanie.

Koszty, ograniczenia i ryzyka lokalnych modeli LLM

Obszar	Osoba indywidualna	NGO
Koszty początkowe	Koszt sprzętu (RAM, dysk).	Koszt sprzętu: zakup lub wydzielenie komputera/serwera (często 32 GB RAM+), koszty serwera zdalnego do hostowania modeli AI
Koszty utrzymania	Koszt działania sprzętu (prąd)	Koszy działania sprzętu (prąd), utrzymania i obsługi.
Koszty pracy	Czas na konfigurację i testowanie modeli.	Czas na konfigurację i testowanie modeli. Czas zespołu na wdrożenie, szkolenia, wsparcie użytkowników.
Jakość modeli	Często niższa niż topowe modele zamknięte.	Często niższa niż topowe modele zamknięte.
Wydajność	Wolniejsze działanie na słabszym sprzęcie.	Wolniejsze działanie na słabszym sprzęcie. Spadki wydajności przy wielu użytkownikach jednocześnie przy prostych serwerach zdalnych.
Skalowalność	Ograniczona, ale możliwe łączenie się zdalnie z serwerem na komputerze.	Trudniejsza rozbudowa (więcej użytkowników lub wykorzystania to konieczność lepszej infrastruktury)
Bezpieczeństwo techniczne	Dane lokalne, ale ryzyko utraty (przy braku backupu)	Ryzyko utraty danych lub nieautoryzowanego dostępu bez rzetelnej konfiguracji.

Pytania dla organizacji

Strategia i cel

- Czy używanie komercyjnych rozwiązań AI jest zgodne z wartościami naszej organizacji i oczekiwaniami naszych interesariuszy?
- Do jakich konkretnych zadań ma służyć lokalny LLM? (podsumowania, tłumaczenia, wyszukiwanie w dokumentach, pisanie, inne?)

- Dlaczego lokalny LLM, a nie gotowe rozwiązania (chmurowe lub API)? Jakie ogranicznie mają te rozwiązania, których szukamy w lokalnym modelu AI (prywatność danych, koszty, opcja offline, kontrola)?
- Kto będzie głównym użytkownikiem: jedna osoba, dział czy cała organizacja?

Dane i prywatność

- Jakie dane będą przetwarzane? Czy zawierają dane osobowe, wrażliwe dane beneficjentów lub dane darczyńców?
- Jakie są nasze zobowiązania prawne (RODO, umowy z darczyńcami, polityki organizacyjne)?
- Czy model będzie miał dostęp do internetu czy tylko do danych lokalnych?
- Kto będzie mieć prawo wglądu w logi i wyniki działania modelu?

Infrastruktura techniczna

- Jakim sprzętem dysponujemy? Ile pamięci RAM, jakie procesory, karty graficzne mają komputery które miałby korzystać z lokalnych modeli?
- Czy mamy własny lub dedykowany serwer lokalny czy tylko komputery robocze?
- Kto w organizacji ma kompetencje do instalacji, konfiguracji i utrzymania systemu?
- Czy mamy plan awaryjny, gdy system przestanie działać?

Zasoby i organizacja

- Ile czasu tygodniowo ktoś może poświęcić na utrzymanie systemu?
- Jaki jest nasz budżet: jednorazowy (sprzęt) i operacyjny (energia, czas osoby technicznej)?
- Czy użytkownicy są gotowi na zmianę sposobu pracy? Co może pomóc im zastąpić korzystanie z rozwiązań komercyjnych lokalnie działającymi modelami AI?
- Kto podejmuje decyzję o wdrożeniu i kto ponosi odpowiedzialność?