

Architektura Empisto: zaawansowane AI w całości na twojej infrastrukturze

15 kwietnia 2026

Nowoczesne AI jest cenne, ale większość tej wartości siedzi za chmurowym API, do którego muszą trafić twoje najbardziej wrażliwe dane. Empisto przyjmuje odwrotne założenie: platforma, modele i twoje dane działają w całości wewnątrz twojej infrastruktury, a do działania systemu nic nie musi jej opuszczać. Ten dokument opisuje, jak to jest zbudowane.

Jedna platforma, wewnątrz twoich murów

Empisto instaluje się jako jedna platforma na sprzęcie, który kontrolujesz, lokalnie lub w prywatnym centrum danych. Wszystko, czego produkt potrzebuje, działa właśnie tam: modele językowe, indeks wyszukiwania zbudowany z twoich dokumentów, aplikacje, które licencjonujesz, oraz środowisko uruchomieniowe dla aplikacji własnych. Nie ma osobnej dzierżawy w chmurze ani żadnego kroku w normalnym użyciu, w którym zapytanie trafiłoby do zewnętrznego modelu.

Granica danych

Linia definiującą architekturę jest granica sieci wokół twojego wdrożenia. Wewnątrz niej znajdują się modele, indeks wyszukiwania, dane aplikacji i logi. Poza nią nie istnieje wychodząca ścieżka dla twoich treści. Platforma jest zbudowana tak, aby działać w pełnej izolacji od sieci, więc może pracować bez żadnego dostępu do internetu. To czyni ją przydatną dla danych regulowanych, materiałów klientów objętych NDA oraz danych z wymogami co do miejsca przechowywania. Tam, gdzie platforma potrzebuje aktualizacji lub licencji, są one obsługiwane jako jawne, możliwe do skontrolowania kroki, a nie jako połączenie w tle, które przy okazji przenosi twoje dane.

Wyszukiwanie przydatne od pierwszego dnia

Pierwszą rzeczą, której zwykle oczekują zespoły, jest zadawanie pytań własnym dokumentom i otrzymywanie ugruntowanych odpowiedzi. Empisto buduje indeks wyszukiwania z treści, którą wskażesz, i od pierwszej instalacji udostępnia nad nią wyszukiwanie i wsparcie. Ponieważ indeks i modele są lokalne, dokumenty stojące za odpowiedzią nigdy nie opuszczają granicy, podobnie jak zapytania wpisywane przez twój zespół.

Aplikacje i SDK

Na wyszukiwaniu opierają się aplikacje, licencjonowane osobno od platformy i wydawane we własnym rytmie. Każda z nich jest samodzielnym wycinkiem: modułem serwerowym, który platforma ładuje, własnymi danymi i własnym miejscem w interfejsie. Gdy dostępne aplikacje nie wystarczają, twoi programiści mogą budować nowe przy pomocy zestawu narzędzi programistycznych (SDK), na tym samym fundamencie i z tą samą gwarancją tego, dokąd mogą trafić dane. Jednostką rozszerzenia jest aplikacja, więc własna implementacja dziedziczy granice platformy, zamiast je obchodzić.

Modele na twoim własnym sprzęcie

Empisto uruchamia otwarte modele na sprzęcie, który należy do ciebie, korzystając ze standardowego oprogramowania do serwowania modeli, więc nie trzeba oddawać możliwości w zamian za kontrolę. Wagi modeli i wnioskowanie działają lokalnie. Nie ma rozliczanego za token wywołania do zewnętrznego dostawcy, co umożliwia model licencyjny o stałym koszcie i sprawia, że użycie nie zamienia się w licznik.

Model komercyjny zgodny z obietnicą

Empisto jest licencjonowane, ze wsparciem i aktualizacjami wersji, a nie rozliczane za token czy za jednostkę przetworzonych danych. To nie jest wyłącznie decyzja cenowa. Ponieważ przychód pochodzi z licencji, nie ma komercyjnego powodu, aby zbierać, przechowywać ani trenować na twoich danych, i architektura to odzwierciedla. Zaufanie ma być weryfikowalne, a nie wypraszone, więc sposób przetwarzania danych i zachowanie sieci są otwarte na niezależny audyt.

Wdrożenie

Typowe wdrożenie umieszcza Empisto w całości w ramach twoich istniejących mechanizmów kontroli sieci, za twoją zaporą i systemami tożsamości. Instalacja odcięta od sieci jest wspierana dla najbardziej rygorystycznych środowisk. Ponieważ wszystko działa w obrębie jednej granicy, jest jeden obwód do przeanalizowania, a nie rozproszony zestaw integracji chmurowych do sprawdzenia.