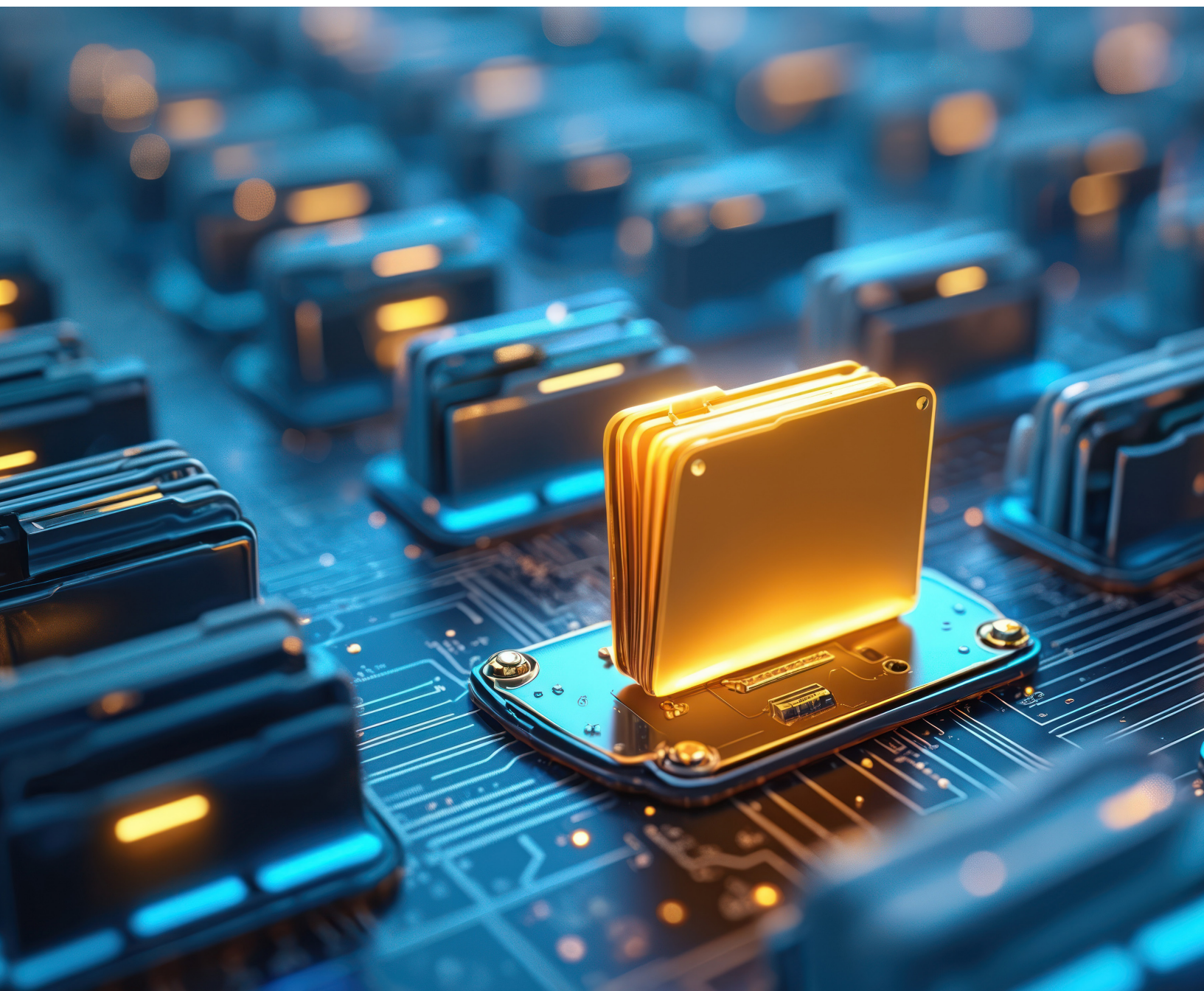


W Maxto ITS łączymy technologie, ludzi, procesy i systemy w spójną oraz bezpieczną całość. Tworzymy rozwiązania IT, które rozwijają biznes i wzmacniają konkurencyjność firm.

MAXTOITS
Future is Possible

Platforma RAG (Retrieval Augmented Generation)



Wdrożenie Platformy AI Retrieval Augmented Generation (RAG) – nowoczesne zarządzanie wiedzą w dużej firmie

W obliczu zarządzania tysiącami dokumentów, regulaminów, umów, ofert czy faktur tradycyjne metody wyszukiwania i analizy danych stają się niewystarczające.

Platformy AI typu Retrieval Augmented Generation (RAG) to zaawansowane rozwiązanie, które łączy możliwości dużych modeli językowych (LLM) z dynamicznym dostępem do wewnętrznych zasobów danych, zapewniając precyzyjne, kontekstowe i skalowalne zarządzanie wiedzą. Niniejszy dokument przedstawia techniczne podstawy RAG, kluczowe korzyści dla organizacji oraz argumenty za jego wdrożeniem.



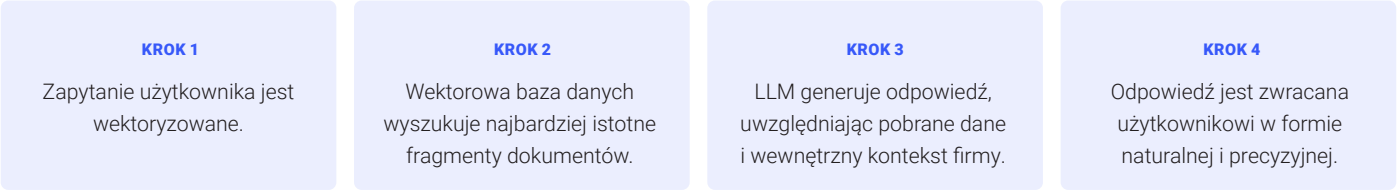
Intel Enterprise RAG – dlaczego ta platforma.

 Rozpoznawalność i zaufanie Intel to 48 lat historii firmy i jeden z największych graczy technologicznych na świecie. Gwarancja stabilności i wysokich standardów.	 Infrastruktura Zapewniona kompatybilność z infrastrukturą serwerową wyposażoną w procesory Intel Xeon.	 Bezpieczeństwo i wsparcie Support 24/7, SLA response poniżej 4h, aktualizacje systemowe, Intel Hardware Security Modules, standard bezpieczeństwa FIPS 140-2.	 Lokalność R&D w Gdańsku. Polski zespół pracujący nad rozwiązaniami software, które Intel dostarcza na rynek światowy.
--	---	--	--

01 | Architektura i działanie systemu RAG

 Model wektoryzacji Dokumenty (np. umowy, regulaminy, raporty) są przekształcane w wektorowe reprezentacje za pomocą modeli językowych. Modele te mapują semantyczne znaczenie tekstu, umożliwiając efektywne wyszukiwanie kontekstowe.	 Wektorowa baza danych Wektory dokumentów są przechowywane w specjalistycznych bazach kontekstowych. Systemy te umożliwiają szybkie wyszukiwanie (w czasie bliskim rzeczywistemu) najbardziej istotnych fragmentów danych na podstawie zapytań użytkowników.
 Duży polski model językowy (LLM)  LLM (np. PLLUM - model dostosowany do języka polskiego) generuje odpowiedzi, łącząc wiedzę z treningu z fragmentami dokumentów pobranymi z bazy wektorowej. Mechanizm ten zapewnia, że odpowiedzi są zgodne z wewnętrznymi regulacjami i specyfiką firmy.	 Mechanizm integracji System RAG integruje dane wejściowe z istniejącymi systemami IT (np. CRM, ERP, SharePoint lub inne) za pomocą API, umożliwiając płynne włączenie do obecnej infrastruktury.

Proces działania



02 | Kluczowe korzyści techniczne

Wysoka precyzja i efektywność wyszukiwania

Platformy AI typu RAG eliminują problem nieefektywnego wyszukiwania w dużych zbiorach danych. Dzięki semantycznemu wyszukiwaniu wektorowemu system odnajduje dokładnie te fragmenty dokumentów, które odpowiadają intencji zapytania, minimalizując szum informacyjny. W porównaniu do tradycyjnych systemów opartych na słowach kluczowych, RAG oferuje dużo wyższą trafność wyników.

Skalowalność

Systemy wektorowe zoptymalizowane pod kątem dużych zbiorów danych, umożliwiając obsługę milionów dokumentów bez utraty wydajności. Architektura RAG jest modularna, co pozwala na łatwe skalowanie w miarę wzrostu liczby danych.

Dynamiczna aktualizacja wiedzy

W odróżnieniu od tradycyjnych modeli LLM, które wymagają kosztownego retreningu przy zmianach danych, RAG umożliwia błyskawiczne dodawanie nowych dokumentów do bazy wektorowej. To kluczowe dla firm, które regularnie aktualizują regulaminy, umowy czy procedury.

Bezpieczeństwo i zgodność

RAG może być wdrożony lokalnie (on-premise), co zapewnia pełną kontrolę nad danymi wrażliwymi, takimi jak umowy czy regulacje wewnętrzne. Szyfrowanie danych i zgodność z przepisami (np. RODO) są wbudowane w system, co minimalizuje ryzyko wycieku informacji.

Optymalizacja dla języka polskiego

Modele wektoryzujące i LLM zoptymalizowane dla języka polskiego (np. PLLUM czy Bielik) zapewniają wysoką jakość analizy i generowania tekstu w języku polskim, co jest kluczowe dla firm operujących na polskim rynku. Oczywiście bazowo pracują także w języku angielskim, czyli w praktyce są dwujęzyczne.



03 | Przykłady zastosowań w firmie



Automatyzacja procesów prawnych

RAG umożliwia szybkie wyszukiwanie klauzul w umowach lub regulaminach, wspierając działy prawne w analizie zgodności i przygotowywaniu dokumentów.



Wsparcie compliance

System pozwala na błyskawiczną weryfikację zgodności procedur z najnowszymi przepisami, redukując ryzyko błędów regulacyjnych.



Efektywna obsługa klienta

Integracja z wewnętrzną bazą wiedzy umożliwia automatyczne generowanie odpowiedzi na pytania klientów, zgodnych z obowiązującymi regulaminami.



Zarządzanie wiedzą wewnętrzną

RAG konsoliduje rozproszone źródła danych (np. SharePoint, bazy PDF) w jedną, łatwo dostępną platformę, przyspieszając onboarding i szkolenia.

04 | Koszty i zwrot z inwestycji (ROI)

Choć wdrożenie RAG wymaga początkowej inwestycji w infrastrukturę i konfigurację, przynosi znaczące oszczędności w średniej i dłuższej perspektywie:

Redukcja czasu pracy

Automatyzacja wyszukiwania i analizy dokumentów zmniejsza obciążenie pracowników.

Minimalizacja błędów

Precyzyjne odpowiedzi zgodne z regulacjami obniżają ryzyko kosztownych pomyłek.

Elastyczność

Możliwość szybkiego dostosowania do nowych przepisów bez retreningu modeli.

Łatwość adaptacji

Możliwość stosowania rozwiązania do wszystkich zakresów działania organizacji opartych na wiedzy i dokumentach.

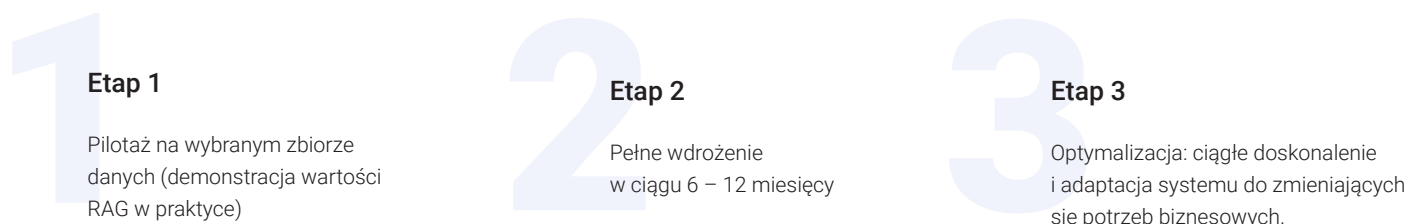
Integracja

Platformy AI można integrować z innymi systemami i osiągać znaczące korzyści przy konsolidacji i wykorzystaniu wielu różnych danych.



05 | Rekomendacja

Wdrożenie Platformy AI typu RAG to strategiczna inwestycja w przyszłość zarządzania wiedzą w firmie. Rozwiązanie to nie tylko zwiększa efektywność operacyjną, ale także buduje przewagę konkurencyjną poprzez szybsze i bardziej precyzyjne reagowanie na potrzeby wewnętrzne i zewnętrzne.



Gwarantujemy usługi na wysokim poziomie

Jakość naszej pracy jest potwierdzona certyfikatami.



Nasi partnerzy technologiczni

Jesteśmy Tytanowym Partnerem Dell Technologies

DELLTechnologies
TITANIUM PARTNER

Lenovo

HUAWEI

VEEAM

CISCO

Alcatel-Lucent

hp

aruba

ENERGY LOGSERVER

ascom

FORTINET

SOPHOS

Microsoft

Lexmark

ZEBRA

vmware by Broadcom

FUJITSU

intel

CLAROTY

**Zapraszamy
do kontaktu**

maxtoits.pl

Maxto ITS Sp. z o.o.
+48 12 266 39 69
info@maxtoits.pl