WE ARE FUTURE | Pomagamy liderom wygrywać w erze AI.

ZAŁĄCZNIK DLA CTO I IT · PAŹDZIERNIK 2026

Architektura **AI Engine**



O DOKUMENCIE

Załącznik opisuje architekturę referencyjną AI Engine: na czym go budujemy, jak działają uprawnienia i tożsamość agentów, jak dobieramy modele i jak spełniamy wymagania bezpieczeństwa. To punkt wyjścia do rozmowy z IT przed uruchomieniem pierwszego procesu.

01. Na czym budujemy AI Engine?

02. Sześć zasad architektury

03. Jak wygląda architektura referencyjna?

04. Uprawnienia, tożsamość i integracje

05. Modele i bramka modeli

06. Bezpieczeństwo i zgodność z regulacjami

07. Od czego zacząć budowę AI Engine?

08. Słownik pojęć technicznych

DLA KOGO

CTO, architekci, zespoły IT i bezpieczeństwa

UZUPEŁNIA

White paper „Jak zbudować firmę przyszłości”



AUTOR

Maciej Rutkowski, CEO · Managing Partner, WeAreFuture

maciej@wearefuture.ai

01 · PLATFORMA

Na czym budujemy **AI Engine?**

Architektura referencyjna AI Engine opiera się na Microsoft Azure, Microsoft 365 i Databricks. Wiele firm korzysta już z tych platform, więc AI Engine powstaje na ich infrastrukturze, a nie obok niej.

INFRASTRUKTURA

Microsoft Azure

Usługi chmurowe w regionie UE: bramka modeli, środowisko agentów, sekrety, monitoring i kopie zapasowe.

TOŻSAMOŚĆ I PRACA

Microsoft 365 i Entra ID

Tożsamość i uprawnienia każdego człowieka i agenta. Poczta, spotkania i dokumenty jako źródła kontekstu.

DANE

Databricks na Azure

Context Lakehouse: dane relacyjne, grafowe i wektorowe w jednym miejscu, z katalogiem danych i uprawnieniami.

AI Engine Context Lakehouse · MCP · Intelligence · Skills · Decisions · Human+AI Teams

Azure jest punktem wyjścia. Dane najbardziej wrażliwe mogą zostać na modelach lokalnych na serwerach firmy (rozdział 05). Gdy firma pracuje na innej chmurze albo potrzebuje pełnego środowiska on-premise, architekturę dobieramy razem z jej zespołem IT.

Co należy do firmy?

Ustalamy, kto administruje środowiskiem oraz gdzie są dane i logi. Dane, procedury, kryteria decyzji i kod napisany dla firmy należą do niej. Z gotowych modułów WeAreFuture firma korzysta na zasadach z umowy.

Między projektami przenosimy metodę, moduły i Skills napisane przez nas. Dane, dokumenty i Decisions firmy zostają w jej AI Engine. Zmiana modelu albo partnera nie oznacza budowy od zera. Pięć elementów zostaje w firmie na lata. Model AI jest wymienny.



02 · ZASADY

Sześć zasad **architektury**

AI Engine nie jest jednym produktem. To zestaw elementów, które buduje się etapami na infrastrukturze, którą firma już ma. Sześć zasad decyduje o tym, czy po trzech latach firma ma spójną platformę, czy kolekcję pilotaży.

01

Najpierw reguła, potem model.

Co da się zapisać regułą, liczy reguła. Model dostaje to, czego regułą opisać się nie da: tekst, wyjątki i ocenę.

02

Ostatnie słowo ma system.

Ceny, sumy i terminy liczą ERP i kod. Model przygotowuje i uzasadnia, ale nie wylicza wartości, za które firma odpowiada.

03

Wiedza w danych, nie w kodzie.

Procedury, kryteria i słowniki są w Skills, Decisions i Context Lakehouse. Wymiana modelu to zmiana konfiguracji i testy jakości, a nie przebudowa.

04

Uprawnienia egzekwuje system.

Dostęp egzekwuje warstwa danych, a nie prompt ani aplikacja. Asystent widzi dokładnie to, co człowiek, bo pyta z tymi samymi uprawnieniami.

05

Człowiek decyduje w ustalonych miejscach.

Punkty akceptacji są częścią przepływu, a nie dodatkiem. Zapis do systemów zewnętrznych domyślnie wymaga zatwierdzenia.

06

Model dobrany do zadania.

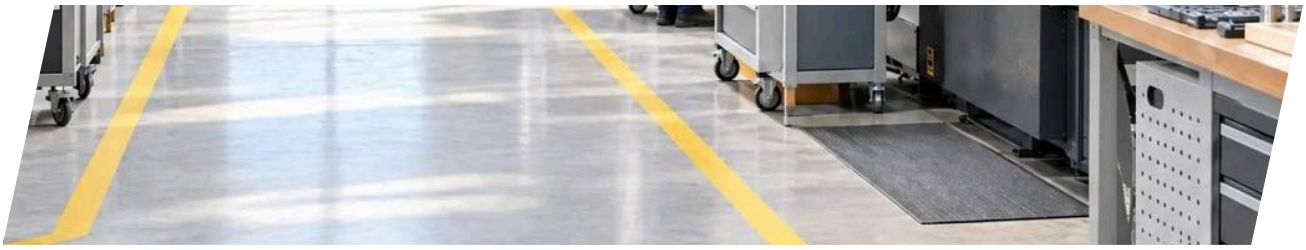
Model wybieramy według jakości, wrażliwości danych, czasu odpowiedzi i kosztu. Nie każde zadanie potrzebuje najmocniejszego modelu.

Tożsamość i ślad w każdym wywołaniu. Każde wywołanie ma tożsamość: człowieka, w którego imieniu działa asystent, albo agenta z własnym kontem. Każde trafia do audytu z kontekstem, wersją procedury i wersją modelu. Bez tożsamości nie ma uprawnień ani audytu.

95%

organizacji nie widziało mierzalnego zwrotu z generatywnej AI. Autorzy badania wskazują przyczyny architektoniczne: kruche przepływy pracy i narzędzia, które nie uczą się z kontekstu firmy.

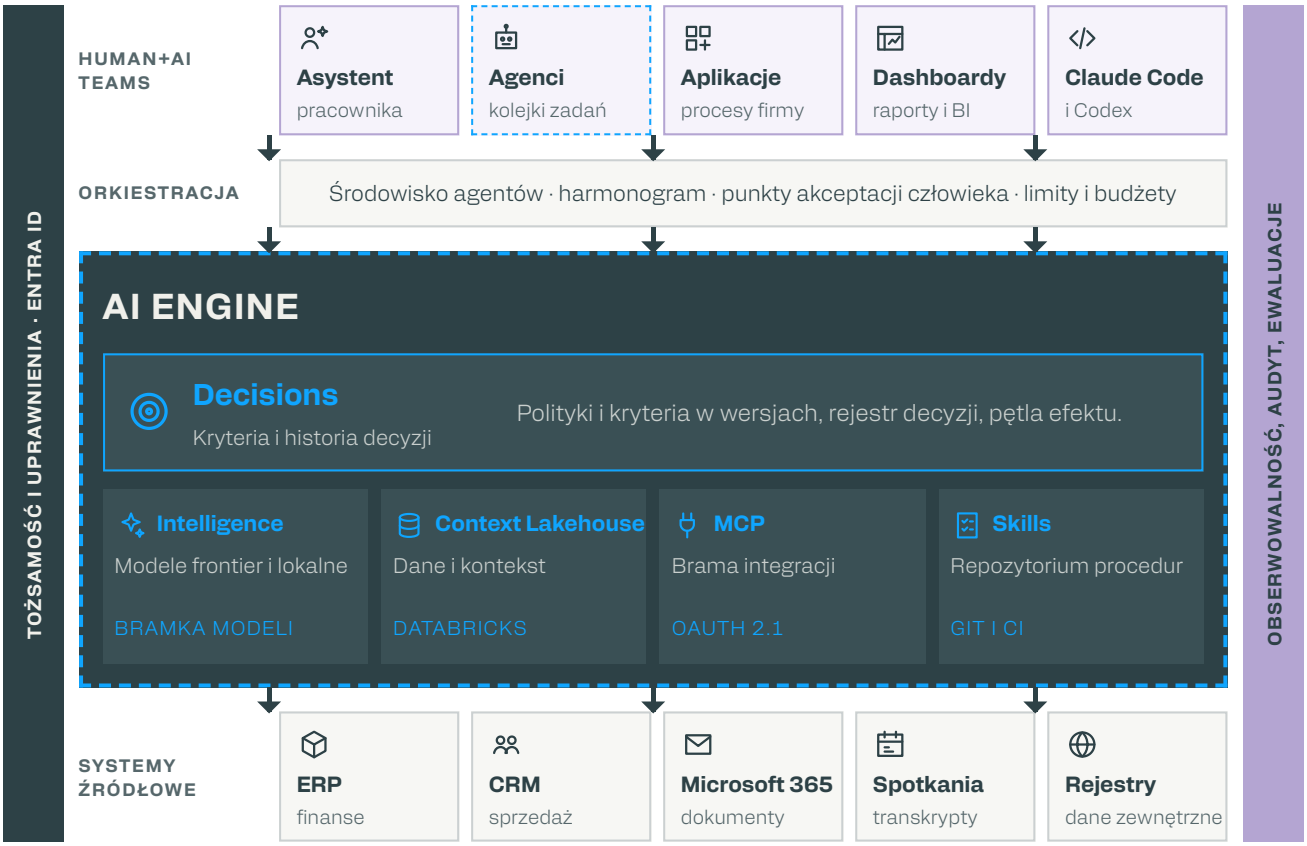
MIT NANDA, The GenAI Divide: State of AI in Business 2025



03 · ARCHITEKTURA

Jak wygląda architektura **referencyjna?**

Sześć elementów AI Engine leży między kanałami, z których korzystają ludzie i agenci, a systemami źródłowymi firmy. Tożsamość oraz obserwowalność przecinają wszystkie warstwy. Ludzie pracują z AI Engine przez asystenta, agentów, aplikacje, dashboardy i raporty oraz narzędzia takie jak Claude Code i Codex, zawsze z tym samym kontekstem i uprawnieniami.



Dane, procedury i decyzje

Context Lakehouse budujemy na Databricks: dane z systemów w warstwach surowych, uporządkowanych i obiektów biznesowych oraz warstwa kontekstu dla modeli. Ludzie korzystają z niego w raportach, a asystenci i agenci przez MCP, zawsze z tymi samymi uprawnieniami. Każde źródło ma właściciela, pochodzenie i wersję.

Skills przechowują procedury, a Decisions kryteria i historię decyzji. Obie warstwy są wersjonowane, więc każdą decyzję agenta można powiązać z wersją procedury i polityki, która wtedy obowiązywała.

> **Uczenie to aktualizacja kontekstu, nie trenowanie modelu.** Zmiana wiedzy lub Skills przechodzi przegląd właściciela, sprawdzenie na przykładach i wersjonowanie. Dopiero zatwierdzona wersja trafia do kolejnych zadań. Modeli nie trenujemy na danych klienta.

04 · INTEGRACJE

Uprawnienia, tożsamość **i integracje**

Model Context Protocol (MCP) to otwarty standard łączenia asystentów i agentów z narzędziami i danymi. W AI Engine brama MCP jest jedynym miejscem, przez które AI dociera do systemów firmy. Dzięki temu uprawnienia, limity i audyt są w jednym miejscu, a nie w każdej integracji osobno.

1	2	3	4	5
Logowanie Użytkownik loguje się kontem firmowym w Entra ID (SSO).	**Token w imieniu** Asystent dostaje krótkotrwały token OAuth wydany dla tej osoby.	**Brama MCP** Weryfikuje token i wykonuje narzędzie z tożsamością użytkownika.	**Uprawnienia w danych** Baza egzekwuje reguły dostępu tak samo jak dla dashboardu.	**Wynik i audyt** Odpowiedź wraca do asystenta, wywołanie trafia do logu.

Asystent i agent

CECHA	ASYSTENT	AGENT AUTONOMICZNY
Tożsamość	Tożsamość użytkownika, token w jego imieniu	Własne konto w Microsoft Entra ID
Uprawnienia	Dziedziczone jeden do jednego	Minimalne, nadane do konkretnego zadania
Odpowiedzialność	Użytkownik	Wskazany właściciel biznesowy

Agent działający na czyjeś zlecenie pobiera z Entra ID token tej osoby, więc nie zrobi więcej, niż wolno zlecającemu. Każdy agent ma limity kosztu i zakresu danych, osobne poświadczenia i możliwość wyłączenia w minutę przez zablokowanie konta.

Integracje i granice działania

Dla ERP, CRM, poczty, Teams i innych systemów opisujemy dozwolone operacje odczytu i zapisu, tożsamość techniczną, limity oraz obsługę błędów. Najpierw ustalamy, które działania wymagają zatwierdzenia człowieka i jak wrócić do pracy ręcznej.

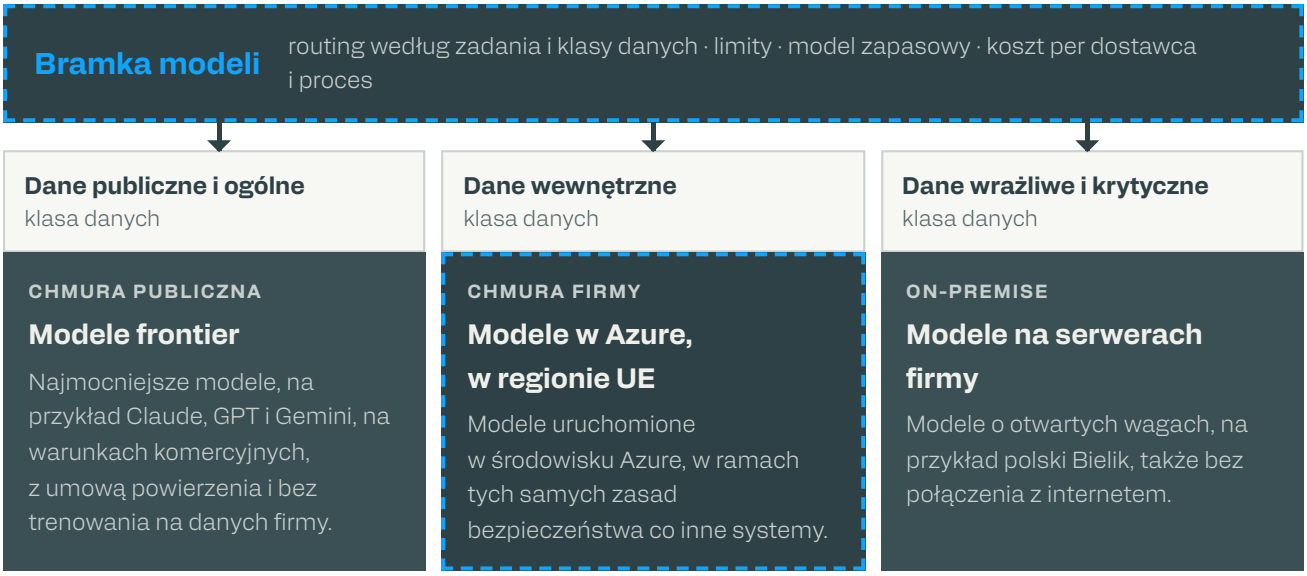
Asystenci i agenci innych dostawców, na przykład Microsoft Copilot albo agent wbudowany w ERP, korzystają z AI Engine przez MCP i konektory, w zakresie uzgodnionym z IT.



05 · INTELLIGENCE

Modele **i bramka modeli**

Wszystkie wywołania modeli przechodzą przez jedną bramkę. Bramka wybiera model według zadania i klasy danych, pilnuje limitów, liczy koszt i przełącza na model zapasowy przy awarii. Aplikacje i agenci nie znają konkretnego dostawcy, znają tylko bramkę.



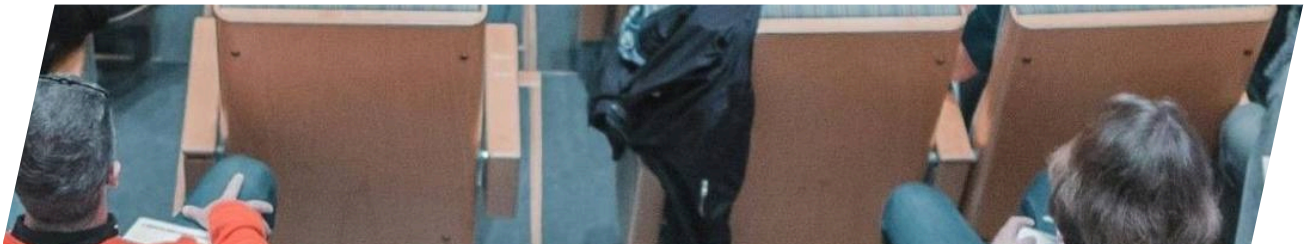
Jak dobieramy modele i liczymy koszt?

Model dobieramy do zadania według jakości, wrażliwości i rezydencji danych, czasu odpowiedzi i kosztu. Dane wrażliwe mogą zostać na modelu lokalnym. Modeli dostawców chmurowych używamy na warunkach komercyjnych, bez trenowania na danych firmy. Zużycie i koszt liczymy per dostawca i per proces. Wymiana modelu nie zmienia danych, Skills, integracji ani kryteriów decyzji.

Kiedy model lokalny?

Model lokalny wchodzi do produkcji dopiero wtedy, gdy na zestawie testów z realnych zadań firmy osiąga jakość ustaloną z właścicielem procesu. Bramka modeli pozwala porównać go z modelem chmurowym na tym samym ruchu, zanim przełączymy proces. Ranking ogólny nie mówi, jak model poradzi sobie z dokumentami konkretnej firmy.

Dobór sprzętu i utrzymanie modeli lokalnych omawiamy osobno, przy konkretnym zastosowaniu.



06 · BEZPIECZEŃSTWO

Bezpieczeństwo **i zgodność z regulacjami**

AI Engine dotyka danych z całej firmy, więc jego bezpieczeństwo projektujemy jak dla systemu krytycznego. Wymagania regulacyjne traktujemy jako założenia architektury.

NIS2

Cyberbezpieczeństwo

Zarządzanie ryzykiem, ciągłość działania, zgłaszanie incydentów i bezpieczeństwo łańcucha dostaw, także dostawców modeli.

AI ACT

Systemy AI

Działania wspierające kompetencje AI osób korzystających z AI (art. 4), klasyfikacja zastosowań według ryzyka, przejrzystość i nadzór człowieka.

RODO

Dane osobowe

Cel i podstawa przetwarzania, ocena skutków dla nowych zastosowań, umowy powierzenia, retencja i prawo do usunięcia.

Zabezpieczenia specyficzne dla AI

- **Ochrona przed wstrzykiwaniem poleceń.** Treści z dokumentów i maili są dla modelu danymi, a nie instrukcjami. Narzędzia z listy dozwolonych, zapis z akceptacją, brak sekretów w kontekście modelu.
- **Rozdzielenie środowisk.** Środowisko deweloperskie działa wyłącznie na danych syntetycznych, z osobnymi kluczami. Skrypty zapisujące dane mają zabezpieczenie przed uruchomieniem na produkcji.
- **Umowy z dostawcami modeli.** Tylko warunki komercyjne z umową powierzenia, bez trenowania na danych firmy, z określonym regionem przetwarzania.

> - **Zgodność i audyt ustalamy przed startem.** Z IT, bezpieczeństwem i prawnikami ustalamy role w przetwarzaniu danych osobowych (umowa powierzenia), kwalifikację zastosowania według AI Act i wymagania NIS2, jeśli firma im podlega. Ustalamy też, jak długo przechowujemy ślad działań ludzi i agentów i jak udostępniamy go audytorowi.



07 · START

Od czego zacząć **budowę AI Engine?**

Fundamenty budujemy na infrastrukturze, którą firma już ma, i od razu pod pierwszy proces z mierzalną wartością. Tak wyglądają pierwsze 90 dni od strony IT. Kolejne procesy korzystają z tych samych elementów.

FUNDAMENTY	PIERWSZY PROCES	WARTOŚĆ
integracja z Entra ID i SSO	Context Lakehouse dla pierwszego procesu	ewaluacje i monitoring jakości
polityka modeli i klas danych	bramka modeli i brama MCP z audytem	raport wartości od punktu odniesienia
wybór pierwszego procesu i miar wartości	asystent i agent z akceptacją człowieka	decyzja o kolejnym procesie

Pytania, na które warto odpowiedzieć przed startem

Gdzie dziś są dane, których potrzebuje pierwszy proces?
Systemy, właściciele, częstotliwość zmian, jakość.

Jak dziś nadawane są uprawnienia?
Role, grupy i wyjątki w Entra ID, które AI musi odwzorować.

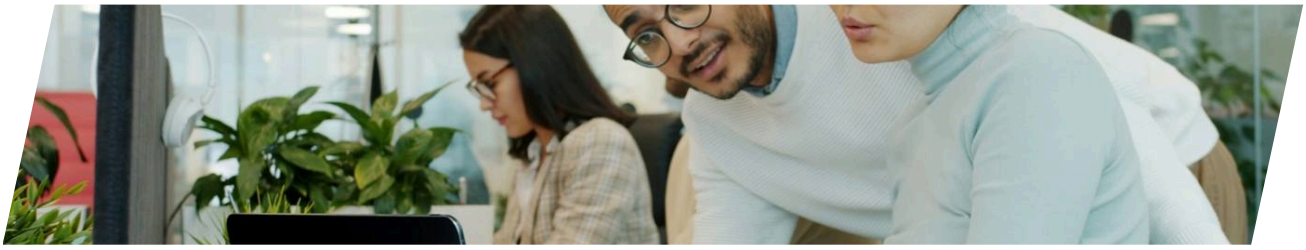
Jak mierzymy dziś proces, który zmieniamy?
Bez punktu odniesienia nie da się pokazać wartości.

Które dane nie mogą opuścić firmy?
Od tego zależy podział między modele chmurowe a lokalne.

Kto będzie właścicielem agentów i procedur?
Odpowiedzialność biznesowa, nie tylko techniczna.

Kto będzie rozwijać AI Engine po uruchomieniu?
Kompetencje, przekazanie wiedzy, wsparcie partnera.

Pełna architektura dla zespołu IT. Szczegółowy model uprawnień, plan wdrożenia i kryteria odbioru przekazujemy zespołowi IT przed warsztatem architektonicznym.



08 · SŁOWNIK

Słownik pojęć **technicznych**

Uzupełnia słownik white papera, w którym opisane są elementy AI Engine i pojęcia biznesowe.

Architektura medalionowa

Podział danych na warstwy: surowe (bronze), uporządkowane (silver) i obiekty biznesowe (gold).

Uprawnienia na poziomie obiektu

Reguły dostępu egzekwowane w bazie dla każdego rekordu, także wobec AI.

OAuth 2.1 i token w imieniu

Standard, w którym asystent dostaje krótkotrwały token użytkownika i działa z jego uprawnieniami.

Bramka modeli

Jedno API do wielu modeli z routingiem według zadania i klasy danych, limitami i rejestrem kosztów. Pozwala zmienić dostawcę bez zmian w reszcie AI Engine.

Tożsamość agenta

Konto agenta w Microsoft Entra ID, z wąskimi uprawnieniami i właścicielem biznesowym. Wyłączenie agenta to zablokowanie konta.

Człowiek w pętli

Punkt w procesie, w którym człowiek przegląda, poprawia lub zatwierdza pracę AI.

Osadzenie (embedding)

Liczbowa reprezentacja znaczenia tekstu, która pozwala wyszukiwać fragmenty podobne treścią.

Klasa danych

Kategoria wrażliwości danych, na przykład ogólnofirmowe, finansowe, kadrowe, od której zależą dostęp i wybór modelu.

Brama MCP

Jedno miejsce, przez które AI dociera do systemów firmy, w otwartym standardzie Model Context Protocol, z uprawnieniami, limitami i audytem.

Model o otwartych wagach

Model, który firma może uruchomić we własnym środowisku, także na serwerach firmy, zamiast korzystać z usługi dostawcy.

Ślad decyzji

Zapis, na jakich danych, według jakich zasad i z czyją akceptacją podjęto decyzję lub wykonano działanie.

ŹRÓDŁA DANYCH

MIT NANDA, The GenAI Divide: State of AI in Business 2025 · AI Act, art. 4 (kompetencje w zakresie AI), w brzmieniu po zmianie przez Digital Omnibus on AI, obowiązującym od 27 lipca 2026

POROZMAWIAJMY O ARCHITEKTURZE

Przejdźmy razem przez waszą infrastrukturę i zaprojektujmy fundamenty AI Engine pod pierwszy proces.

Warsztat architektoniczny z zespołem IT trwa jeden dzień. Wychodzimy z niego z mapą danych i uprawnień, klasami danych oraz planem pierwszych 90 dni.

Pomagamy liderom **wygrywać w erze AI.**



KONTAKT

Napisz do nas, a umówimy warsztat z waszym zespołem IT i bezpieczeństwa.

STRONA
wearefuture.ai

NAPISZ DO NAS
hello@wearefuture.ai

LINKEDIN
linkedin.com/company/wearefuturepl