

ZAPYTANIE O INFORMACJĘ (RFI)

ROZEEZNANIE RYNKU I POZYSKANIE ORIENTACYJNEJ WYCENY

Moduł AI Lean

Lean Healthcare Management

Model LLM o wysokiej zdolności rozumowania + RAG + adaptacja domenowa + kompletna infrastruktura on-premise

Jeden kompletny wariant | jeden model/checkpoint | jedna konfiguracja infrastruktury | jedna wycena

Zamawiający	Państwowy Instytut Medyczny MSWiA
Adres	ul. Wołoska 137, 02-507 Warszawa
Numer RFI	RFI/AI-LEAN/[NRJ]/2026
Data publikacji	7 sierpnia 2026 r.
Termin pytań	11.08.2026
Termin odpowiedzi	13.08.2026
Adres do przestania odpowiedzi	adam.kaniuk@pimmswia.gov.pl

Wersja dokumentu: 1.2 | język odpowiedzi: polski | waluta: PLN

Dokument samodzielny do RFI. Wszystkie istotne założenia zakresowe, ilościowe, jakościowe, techniczne i odbiorowe przyjmowane do wyceny podano w niniejszym dokumencie.

1. Status, cel i zasady RFI

Najważniejsze: RFI służy wyłącznie rozpoznaniu rynku, potwierdzeniu wykonalności i pozyskaniu orientacyjnego oszacowania wartości zamówienia. OPZ nie będzie przekazywany respondentom razem z RFI. Respondent wycenia jeden kompletny wariant na podstawie wszystkich założeń zawartych w niniejszym dokumencie.

- Niniejszy dokument nie stanowi ogłoszenia o zamówieniu, zaproszenia do składania ofert, zaproszenia do negocjacji ani zobowiązania do zawarcia umowy.
- Udział w RFI jest dobrowolny i nieodpłatny. Brak odpowiedzi nie ogranicza możliwości udziału w przyszłym postępowaniu, a udzielenie odpowiedzi nie zapewnia pierwszeństwa ani preferencji.
- Zamawiający może wykorzystać informacje z RFI do oszacowania wartości zamówienia, oceny wykonalności, doprecyzowania SWZ i wzoru umowy oraz zaplanowania finansowania.
- Zamawiający może poprosić o wyjaśnienia lub krótkie spotkanie techniczne. Nie oznacza to negocjacji handlowych.
- Informacje stanowiące tajemnicę przedsiębiorstwa należy wydzielić, oznaczyć i uzasadnić zgodnie z obowiązującymi przepisami.

1.1. Jeden wariant i jedna wycena

- Respondent wskazuje jedną architekturę, jeden bazowy model docelowy, dokładny checkpoint, jedną docelową kwantyzację i jedną kompletną konfigurację infrastruktury on-premise.
- Nie należy przedstawiać alternatywnych wariantów cenowych, wariantów ograniczonych, zamienników, praw opcji ani osobnych cen dla rozwiązań dodatkowych.
- Porównanie co najmniej dwóch konfiguracji kandydackich, wymagane w przyszłym postępowaniu jako element raportu benchmarkowego, należy uwzględnić w koszcie realizacji, ale nie stanowi ono drugiego wariantu cenowego RFI.
- Cena obejmuje wszystkie składniki niezbędne do spełnienia wymagań, także przewody, połączenia, elementy montażowe, licencje, narzędzia utrzymaniowe i usługi, których nie wymieniono osobno, ale są potrzebne do kompletnego działania.
- Koszty należy podać wyłącznie jako: (1) Sprzęt IT, (2) Pozostały zakres zamówienia oraz (3) Razem.

1.2. Status wymagań M, W i O

Status	Znaczenie na potrzeby wyceny RFI
M - obowiązkowe	Musi zostać spełnione i ujęte w jednej cenie. Niespełnienie oznacza odstępstwo od oczekiwanego zakresu.
W - wysoko pożądane	Nie jest obowiązkowe w cenie podstawowej. Respondent wskazuje TAK/NIE, czy funkcja jest zawarta w proponowanej konfiguracji bez osobnej wyceny. Zadeklarowane spełnienie staje się elementem opisu wariantu.
O - opcjonalne	Nie jest objęte obowiązkową ceną. RFI nie przewiduje osobnej ceny opcji; respondent może jedynie wskazać, czy funkcja jest zawarta w cenie bez dopłaty.

Status	Wymagania nieobowiązkowe, które należy ujawnić
W	Wybór poziomu odpowiedzi: skrócona, ekspercka lub dydaktyczna; natywny kontekst modelu $\geq 262\,144$ tokenów; pomocnicza ocena UAT przez co najmniej dwa modele sędziowskie; SSO/OIDC/SAML; powtarzalne pakiety wdrożeniowe z deklaracyjną konfiguracją; pomiar energii Wh/1000 tokenów.
O	Efektywny kontekst $\geq 1\,000\,000$ tokenów. Funkcja poza obowiązkowym zakresem i bez odrębnej ceny w RFI.

2. Kontekst projektu i przedmiot planowanego zamówienia

Planowane zamówienie dotyczy zaprojektowania, wykonania, dostawy i wdrożenia kompletnego modułu AI Lean wspierającego analizę, interpretację i wizualizację procesów zarządczych w jednostkach ochrony zdrowia. Moduł będzie wykorzystywany w środowisku dydaktycznym i warsztatowym Lean Healthcare Management oraz będzie cyfrowym symulatorem i platformą szkoleniową.

Rozwiązanie ma łączyć bazowy model językowy o wysokiej zdolności rozumowania, adaptację domenową, lokalną bazę wiedzy RAG, bibliotekę walidowanych promptów, aplikację użytkownika i administratora, integrację, mechanizmy bezpieczeństwa oraz kompletną infrastrukturę on-premise. Wszystkie komponenty AI i dokumentowe mają działać lokalnie, bez przekazywania danych do zewnętrznych usług.

Zakres niekliniczny: System ma charakter dydaktyczny, analityczny i zarządczy. Nie służy diagnozowaniu, leczeniu, interpretacji wyników badań ani podejmowaniu decyzji klinicznych lub personalnych. Domyślnie wykorzystywane są dane syntetyczne, dydaktyczne lub zanonimizowane.

2.1. Oczekiwane rezultaty

Obszar	Oczekiwany rezultat
Asystent AI Lean	Obsługa języka polskiego oraz terminologii i metodyki Lean Healthcare Management.
Analiza procesów	Identyfikacja etapów, ról, przepływu, oczekiwania, marnotrawstwa, wąskich gardeł, zależności i przyczyn źródłowych.
Metody Lean	Wsparcie 5 Why, Ishikawy, raportów A3, VSM oraz projektowania i porównywania wariantów usprawnień.
Symulator i dydaktyka	Interpretacja wyników symulacji, rekomendowanie dalszych eksperymentów oraz tryb pracy trenera i uczestnika.
RAG i źródła	Odpowiedzi oparte na zatwierdzonych materiałach, cytowanie źródeł, rozdzielanie faktów, hipotez i rekomendacji oraz prawidłowa odmowa przy braku podstaw.
Raportowanie	Generowanie edytowalnych raportów, danych strukturalnych, eksportów i podsumowań sesji.
Gotowość operacyjna	Kompletna infrastruktura, bezpieczeństwo, testy, dokumentacja, szkolenia, kod źródłowy, prawa, gwarancja i serwis.

3. Pełny zakres objęty jedną ceną

Respondent przyjmuje odpowiedzialność za kompletny, kompatybilny i możliwy do eksploatacji rezultat. Poniższy zakres należy ująć w jednej wycenie, bez wyłączania elementów pomocniczych koniecznych do osiągnięcia parametrów opisanych w dalszych rozdziałach.

Lp.	Obszar	Zakres do uwzględnienia w cenie
1	Analiza, architektura i sizing	Doprecyzowanie przypadków użycia, architektury logicznej i technicznej, przepływów danych, modelu bezpieczeństwa, wymagań fizycznych, sizingu oraz sposobu integracji.
2	Kompletna infrastruktura on-premise	Dostawa i instalacja wszystkich zasobów obliczeniowych, pamięciowych, sieciowych, pamięci masowej, montażowych, zasilających, systemowych, licencyjnych, monitoringowych i backupowych. Pamięć przeznaczona na obsługę modelu LLM co najmniej 384 GB
3	Bazowy model docelowy	Dobór i lokalne uruchomienie modelu o wysokiej zdolności rozumowania, dokładnego checkpointu i docelowej kwantyzacji, wraz z dowodami benchmarkowymi, manifestem i skryptami.
4	Adaptacja domenowa i dane	Przygotowanie korpusu instrukcyjnego, zbiorów treningowych i walidacyjnych, data card oraz wykonanie uzasadnionej, odtwarzalnej adaptacji modelu; rezultat w postaci wag po dostrojeniu albo adaptera domenowego.
5	RAG i pipeline dokumentowy	Repozytorium wiedzy, import i analiza dokumentów, OCR/VLM, metadane, lokalne embeddingi, wyszukiwanie hybrydowe, reranking, cytowania, wersjonowanie i kontrola publikacji.
6	Biblioteka promptów	Projektowanie, walidacja, wersjonowanie, zatwierdzanie i wycofywanie minimum 25 szablonów dla kluczowych scenariuszy Lean.
7	Aplikacja, panel i API	Interfejs użytkownika, panel administratora, sesje, role, historia, informacja zwrotna, raportowanie, eksport oraz dokumentowane API REST/JSON.
8	Integracje	Integracja z cyfrowym symulatorem, platformą szkoleniową, repozytorium dokumentów, lokalnymi endpointami modeli, monitoringiem/SIEM i - jako wymaganie W - SSO.
9	Bezpieczeństwo i zgodność	Pełna lokalność, szyfrowanie, RBAC, audyt, backup i odtworzenie, prompt injection, nadzór człowieka, AI Act, RODO, prawo autorskie, WCAG i praca offline.
10	Testy, UAT i produkcja	Benchmarki, testy funkcjonalne, domenowe Lean, RAG, dokumentów, bezpieczeństwa, wydajności, integracji, infrastruktury i stabilności 72 h oraz konwersja środowiska do produkcji.
11	Dokumentacja i szkolenia	Dokumentacja techniczna, powykonawcza, użytkowa, administracyjna i developerska, model card, data card, instrukcje utrzymania oraz szkolenia wszystkich grup.
12	Kod, prawa, gwarancja i serwis	Pakiet kodu źródłowego i utrzymaniowego, SBOM, build/deploy offline, prawa do rozwoju przez podmiot trzeci, gwarancja infrastruktury i oprogramowania oraz obsługa błędów.

4. Wymagania funkcjonalne i użytkowe

Obszar	Założenia obowiązkowe do wyceny
Język i wejścia	Naturalny język polski i terminologia Lean Healthcare Management. Wejście: opis procesu w tekście oraz dane z plików i integracji w uzgodnionych formatach.
Analiza procesu	Identyfikacja etapów, ról, wejść, wyjść, zależności, oczekiwania, przepływu i potencjalnych problemów organizacyjnych.
Marnotrawstwo i ograniczenia	Klasyfikacja marnotrawstwa z uzasadnieniem i źródłem; wskazanie potencjalnych wąskich gardeł, ograniczeń przepustowości i hipotez wymagających weryfikacji danymi.
Przyczyny źródłowe	Wsparcie 5 Why, diagramu Ishikawy lub podejścia równoważnego, z wyraźnym oddzieleniem faktów od hipotez.
A3 i VSM	Generowanie edytowalnej struktury A3: problem, stan obecny, analiza, cel, działania, miary i plan kontroli. Strukturalny opis VSM/mapy procesu możliwy do wizualizacji w symulatorze lub eksporcie.
Warianty usprawnień	Co najmniej trzy warianty wraz z założeniami, ryzykami i spodziewanym wpływem. Porównanie co najmniej według czasu, jakości, obciążenia personelu, wykorzystania zasobów, kosztu i ryzyka.
Symulator	Interpretacja wyników symulacji, wskazanie różnic względem wariantu bazowego i rekomendowanie dalszych eksperymentów.
Tryb dydaktyczny	Wsparcie prowadzącego i uczestnika przez pytania naprowadzające, wskazówki, podsumowania i informację zwrotną.
Raportowanie	Podsumowania sesji, raporty warsztatowe i rekomendacje w formie edytowalnej.
Ugruntowanie odpowiedzi	Cytowania dokumentu, rozdziału i - jeżeli dostępne - strony; rozdzielanie treści źródłowych, obliczeń, założeń, hipotez i rekomendacji; pytanie uzupełniające lub odmowa przy braku podstaw.
Prompty	Minimum 25 walidowanych szablonów; uprawniony użytkownik może tworzyć, testować, wersjonować, zatwierdzać i wycofywać prompty.
Sesje i doskonalenie	Historia sesji obejmuje kontekst, źródła i wersję modelu. Użytkownik może ocenić odpowiedź, oznaczyć błąd i przekazać korektę do kontrolowanego procesu; brak automatycznego uczenia.
Role	Co najmniej: administrator AI, ekspert Lean, trener, uczestnik i audytor.
Eksport	Co najmniej DOCX, PDF i JSON; dane tabelaryczne również CSV/XLSX.
Dokumenty i grafika	Odczyt PDF tekstowych i skanowanych, tabel, wykresów, schematów i grafik za pomocą lokalnych komponentów.
Ograniczenie kliniczne	Brak porad klinicznych i decyzji medycznych; stałe komunikowanie dydaktycznego i organizacyjnego zakresu systemu.

5. Model LLM, benchmarki i adaptacja domenowa

Dopuszczalna forma adaptacji: Adaptacja domenowa jest elementem obowiązkowym, lecz Zamawiający nie narzuca techniki. Wykonawca może zastosować dostrojenie wag, PEFT, adapter domenowy albo rozwiązanie równoważne, adekwatne do licencji, zasobów i celu. Odbieranym rezultatem mogą być wagi po dostrojeniu lub adapter wraz z tokenizerem, konfiguracją, model card i instrukcją uruchomienia. RAG pozostaje podstawowym źródłem wiedzy faktograficznej.

Obszar	Założenie przyjmowane do wyceny
Neutralność technologiczna	Brak wskazania konkretnego producenta, modelu, architektury lub minimalnej liczby parametrów. Respondent podaje jedną konfigurację spełniającą wszystkie bramki.
Porównanie kandydatów	W koscie należy uwzględnić raport porównujący co najmniej dwie dokładnie oznaczone konfiguracje kandydackie w ich docelowych kwantyzacjach, a następnie wskazanie jednej konfiguracji produkcyjnej. W RFI nadal obowiązuje jedna cena i jeden wariant.
Identyfikacja	Nazwa, wersja, suma kontrolna, licencja, sposób pozyskania, format wag, kwantyzacja, mapa precyzji, rozmiar checkpointu, tokenizer, szablon rozmowy, ustawienia rozumowania, parametry generacji i konfiguracja uruchomieniowa.
Profil benchmarkowy i produkcyjny	Odrębna dokumentacja profilu benchmarkowego oraz produkcyjnego, obejmującego prompt systemowy Lean, RAG, adapter, guardrails, routing i integracje.
Progi inteligencji modelu	HLE text-only >=35,0% dla rodziny modelu. Dokładny checkpoint w docelowej kwantyzacji: GPQA-Diamond >=89,0% oraz MMLU-Pro >=81,0% albo AIME 2026 >=93,0% bez narzędzi.
Metodyka benchmarków	Pass@1; niezmienione oficjalne lub legalnie autoryzowane zbiory; bez adaptacji domenowej, adapterów, RAG, Internetu, wyszukiwarki, kalkulatora, interpretera kodu i dodatkowego modelu, chyba że oficjalna metodologia stanowi inaczej.
Dowody	Dopuszczalny wynik publiczny z katalogu SWZ, równoważny przedmiotowy środek dowodowy albo podpisany raport pełnego testu Wykonawcy wraz ze skryptami, manifestem, surowymi odpowiedziami i logami. Możliwe zdalne potwierdzenie na środowisku Wykonawcy.

Obszar	Założenie przyjmowane do wyceny
Kwantyzacja	Co najmniej 4-bitowa albo precyzja wyższa. Bazowe tensory podlegające kwantyzacji nie mogą mieć nominalnej precyzji poniżej 4 bitów. Dopuszczalna precyzja mieszana i wyższa precyzja warstw wrażliwych.
Retencja jakości	Porównanie z odpowiednim wariantem tego samego modelu w wyższej precyzji na co najmniej trzech benchmarkach, w tym GPQA-Diamond. Średnia retencja $\geq 97,5\%$; spadek w każdym benchmarku $\leq 3,0$ p.p.
Kontekst	Obowiązkowo efektywny i przetestowany kontekst $\geq 128\ 000$ tokenów po uwzględnieniu promptu systemowego, narzędzi i RAG. Natywny kontekst $\geq 262\ 144$ tokenów - W. Efektywny kontekst $\geq 1\ 000\ 000$ tokenów - O.
Narzędzia i JSON	Wywoływanie narzędzi lub mechanizm równoważny, stabilny JSON zgodny ze schematem, streaming oraz sterowanie głębokością rozumowania, długością i parametrami generacji.
Parametry modelu	Obowiązkowe ujawnienie parametrów całkowitych i aktywnych oraz opis architektury dense/MoE/innej; brak minimalnego progu parametrów.
Modele pomocnicze	Dopuszczalne do OCR, dokumentów, grafiki, ekstrakcji, routingu i prostych zadań. Nie mogą zastąpić głównego modelu Lean, chyba że samodzielnie spełnią wszystkie bramki.
Zbiór treningowy	Do wyceny przyjąć co najmniej 3 000 zwalidowanych przykładów. Mniejszy zbiór jest dopuszczalny wyłącznie po wykazaniu wszystkich metryk odbiorowych i zaakceptowaniu uzasadnienia jakościowego.
Zakres przykładów	Pytania-odpowiedzi, marnotrawstwo, wąskie gardła, 5 Why, A3, VSM, studia przypadków, scenariusze warsztatowe i interpretacja symulacji.
Rozdzielenie danych	Zbiory treningowe, walidacyjne i scenariusze UAT muszą być rozłączne. UAT nie może służyć do strojenia modelu, promptów, RAG, rerankera, guardrails ani wyboru checkpointu.
Przygotowanie materiału	Niedopuszczalne jest mechaniczne podanie całej książki jako jedynej metody treningu. Materiał musi zostać przekształcony w uporządkowane, ekspercko zatwierdzone przykłady instrukcyjne.
Prawa i dane	Wyłącznie materiały z prawem wykorzystania oraz dane syntetyczne, dydaktyczne lub skutecznie zanonimizowane.
Reprodukowalność	Konfiguracja adaptacji i testów, wersje bibliotek, seedy, metryki, logi i artefakty umożliwiające powtórzenie procesu.
Wyjaśnienia	Interfejs nie ujawnia surowego wewnętrznego toku rozumowania; użytkownik otrzymuje zwięzłe uzasadnienie, źródła i jawne założenia.
Przenośność	Kontrolowana zmiana modelu bez utraty danych RAG, promptów i historii testów; stabilna lokalna warstwa adaptera modelowego.

6. RAG, dane dokumentowe i biblioteka promptów

Skala danych RAG: OPZ nie ustala odrębnej minimalnej liczby dokumentów ani liczby GB korpusu startowego. Parametrem porównywalnym do wyceny jest indeks obejmujący co najmniej 100 000 fragmentów wiedzy oraz wyszukiwanie hybrydowe z rerankingiem w czasie $p95 \leq 3$ s. Respondent ma podać własne założenia sizingowe dotyczące liczby dokumentów, stron i GB, ale bez osobnej pozycji cenowej.

Obszar	Wymagania obowiązkowe
Zakres wiedzy	Książka „Lean Healthcare Management”, kolejne zatwierdzone wydania, materiały szkoleniowe, słownik pojęć, studia przypadków, wzorcowe A3, mapy procesów, dokumentacja symulatora i materiały eksperckie zatwierdzone przez Zamawiającego.
Skala i wydajność	Indeks $\geq 100\ 000$ fragmentów; wyszukiwanie hybrydowe wraz z rerankingiem $p95 \leq 3$ s, bez czasu generowania odpowiedzi.
Formaty	PDF, DOCX, PPTX, TXT, HTML, XLSX, CSV, JPG i PNG; kontrolowany OCR dla skanów.
Pipeline dokumentowy	Rozpoznanie PDF z tekstem, skanu, tabeli, wykresu i schematu oraz kierowanie do właściwego lokalnego parsera/OCR/VLM.
Metadane fragmentów	Tytuł, wersja, autor/źródło, data, rozdział, strona, typ treści i status zatwierdzenia - jeżeli dostępne.
Retrieval	Wyszukiwanie semantyczne i słownikowe oraz reranking lub mechanizm równoważny; Recall@5 $\geq 0,85$.
Cytowania	Odwołanie do źródła i podgląd właściwego fragmentu dokumentu lub obrazu strony; poprawność cytowania $\geq 90\%$, pokrycie cytowaniem $\geq 95\%$.
Brak źródła	Zgłoszenie braku informacji i odmowa tworzenia odpowiedzi spekulatywnej; poprawna reakcja w $\geq 95\%$ przypadków testowych.
Publikacja	Dodanie, aktualizacja i wycofanie dokumentu po zatwierdzeniu przez uprawnionego eksperta lub administratora.
Wersjonowanie	Historia wersji dokumentów, indeksów, embeddingów i modeli dokumentowych oraz możliwość odtworzenia stanu użytego w konkretnej odpowiedzi.

Uprawnienia	Wyniki wyszukiwania respektują dostęp do dokumentów i nie ujawniają treści niedostępnych dla użytkownika.
Lokalność	Embeddingi, reranker, OCR/VLM, baza wektorowa i wszystkie komponenty dokumentowe działają lokalnie; brak zewnętrznych usług.
Bezpieczeństwo dokumentów	Wykrywanie i neutralizacja prompt injection w dokumentach oraz ograniczanie nieuprawnionej eksfiltracji treści.
Aktualizacja wiedzy	Aktualizacja RAG nie wymaga ponownego trenowania modelu; kolejny fine-tuning wyłącznie po uzasadnieniu.
Prawo autorskie	Ograniczenie nadmiernego reprodukowania źródeł; preferowane streszczenia, krótkie cytaty i wskazanie źródła.
Prompty	Minimum 25 szablonów. Każdy ma właściciela, wersję, zakres, dane testowe, kryteria poprawności i status zatwierdzenia.
Zakres promptów	Analiza procesu, marnotrawstwo, wąskie gardła, 5 Why/Ishikawa, A3, VSM, warianty usprawnień, wyniki symulacji, tryb warsztatowy i raport końcowy.

7. Aplikacja, API i integracje

Obszar	Wymaganie
Interfejs użytkownika	Aplikacja webowa z obsługą sesji, historii, źródeł, feedbacku, raportowania i eksportów; responsywna i zgodna z WCAG 2.1 AA.
Panel administracyjny	Zarządzanie modelami i wersjami, RAG, dokumentami, promptami, rolami, retencją, logami, bezpieczeństwem i wyłączeniem modelu.
API	Dokumentowane REST/JSON zgodne z OpenAPI lub rozwiązanie równoważne; jeden stabilny lokalny interfejs usługi niezależnie od architektury wewnętrznej.
Symulator	Pobieranie danych scenariusza i wyników oraz zwracanie ustrukturyzowanych wniosków, rekomendacji i danych do wizualizacji.
Platforma szkoleniowa	Uruchomienie sesji AI z poziomu platformy, przekazanie kontekstu ćwiczenia i zapis podsumowania.
Repozytorium dokumentów	Import dokumentów i synchronizacja metadanych z repozytorium wskazanym przez Zamawiającego.
Adapter modelowy	Wyłącznie lokalne endpointy on-premise; kontrolowana wymiana modelu bez przebudowy interfejsu, RAG i integracji.
Monitoring/SIEM	Przekazywanie metryk, alertów i logów bezpieczeństwa do narzędzi Zamawiającego.
Router multimodalny	Kierowanie tekstu, dokumentów, skanów i obrazów do właściwych komponentów i budowa ustrukturyzowanego kontekstu dla głównego LLM.
SSO - W	Integracja z OIDC/SAML lub rozwiązaniem równoważnym i mapowanie ról; respondent wskazuje, czy funkcja jest zawarta w jednej cenie.

8. Infrastruktura on-premise, pojemność i wydajność

Odpowiedzialność Wykonawcy: Zamawiający nie narzuca producenta, liczby urządzeń, topologii, sposobu parallelizmu, organizacji pamięci, systemu operacyjnego, konteneryzacji ani silnika inferencyjnego. Wykonawca odpowiada za sizing, kompletność, zgodność komponentów oraz osiągnięcie wszystkich parametrów na zaoferowanym środowisku.

Obszar	Minimum przyjmowane do wyceny
Lokalizacja	Wszystkie elementy fizycznie u Zamawiającego. Brak chmury, hostowanego inferowania, zewnętrznego API AI i awaryjnego przekierowania.
Zakres komponentów lokalnych	Główny LLM, modele pomocnicze, embeddingi, reranker, OCR, VLM, guardrails, baza wektorowa, router, monitoring i logowanie.
Pamięć inferencyjna	>=384 GB łącznej nominalnej pamięci obliczeniowej faktycznie dostępnej procesom inferencyjnym albo więcej według sizingu; podczas S2 >=5% rezerwy; brak OOM.
Niedopuszczalne mechanizmy	Pamięć masowa, swap, zRAM i awaryjne stronicowanie nie są zaliczane do minimum i nie mogą służyć do spełnienia parametrów.
Pamięć masowa	Miejsce dla co najmniej dwóch wersji modelu, adapterów, indeksów, danych roboczych, logów i kopii; co najmniej 30% wolnej rezerwy po odbiorze.
Sieć i parallelism	Połączenia dobrane do modelu i zastosowanego tensor/pipeline/expert parallelizmu lub rozwiązania równoważnego tak, aby spełnić wydajność i stabilność.
Kompletność	Wszystkie urządzenia, połączenia, kable, elementy sieciowe, montażowe, zasilające, storage, sterowniki, licencje, monitoring, backup i narzędzia utrzymaniowe.
Jedno środowisko	Jedno fizyczne środowisko: najpierw testowo-odbiorowe, następnie po oczyszczeniu danych, rotacji sekretów, hardeningu, skanie, backupie i odtworzeniu - produkcyjne. Bez stałego pre-produkcji.

Dostępność	>=99,5% miesięcznie od odbioru produkcyjnego, z wyłączeniem uzgodnionych okien serwisowych.
Backup i DR	Kopie bazy wiedzy, konfiguracji, promptów, adapterów i metadanych; RPO <=24 h, RTO <=8 h.
Aktualizacje	Podpisane pakiety offline wprowadzane przez kontrolowane repozytorium Zamawiającego; brak obowiązkowej aktywacji i okresowych kontroli online.
Parametry fizyczne	Wykonawca podaje liczbę/formę urządzeń, miejsce, masę, moc typową i maksymalną, zasilanie, chłodzenie, temperaturę, wilgotność, hałas i interfejsy.
Sizing	Rozmiar wag, narzut formatu, KV cache, bufor, kontekst, współbieżność, usługi pomocnicze, storage, rezerwa, pobór mocy i chłodzenie.
Skalowalność	Możliwość zwiększenia wydajności i liczby użytkowników bez zmiany logiki biznesowej, RAG i interfejsów.

8.1. Profile wydajności

Profil	Obciążenie	Wymagane parametry
S1	1 sesja; wejście 4 000 tokenów; wyjście 300 tokenów; bez RAG	TTFT p95 <=5 s; mediana >=20 tokenów/s; pełna odpowiedź p95 <=25 s.
S2	10 sesji równoległych; standardowy profil produkcyjny	TTFT p95 <=10 s; pełna odpowiedź p95 <=60 s; zagregowana wydajność >=60 tokenów/s; rezerwa pamięci >=5%.
Profil rozszerzonego rozumowania	Dopuszczalny dodatkowo	Może mieć dłuższy czas, ale nie zastępuje profilu standardowego w testach S1/S2.

8.2. Stabilność

- Test 72 godziny po rozgrzewce, z cyklicznym obciążeniem S1, S2 i zmiennym, na końcowym modelu, adaptacji, RAG i infrastrukturze.
- 0 nieplanowanych restartów, OOM, zakleszczeń, utraty funkcji krytycznej i zewnętrznych wywołań AI.
- Skuteczność żądań co najmniej 99,5% globalnie i co najmniej 98,0% w każdym 15-minutowym oknie.
- Co najmniej 5% rezerwy pamięci podczas S2 oraz parametry temperatury i mocy w granicach producentów.

9. Bezpieczeństwo, zgodność, prawa i dostępność

Obszar	Wymagania do uwzględnienia
Szyfrowanie	TLS 1.2 lub wyżej w transmisji; szyfrowanie danych w spoczynku zgodnie z polityką Zamawiającego.
RBAC	Zasada minimalnych uprawnień dla modeli, dokumentów, promptów, logów i funkcji administracyjnych.
Audyt	Rejestr użytkownika, czasu, modelu, checkpointu, wersji promptu, źródeł RAG, operacji administracyjnych i wyniku kontroli bezpieczeństwa.
Brak auto-uczenia	Rozmowy i feedback nie aktualizują automatycznie modelu ani bazy treningowej.
Zakaz transferu	Prompty, dokumenty, dane użytkowników, embeddingi, obrazy, logi, odpowiedzi, metadane i artefakty pośrednie nie opuszczają środowiska.
Retencja	Konfigurowalne okresy retencji rozmów, logów i plików oraz bezpieczne usuwanie.
Prompt injection	Ochrona instrukcji systemowych, wykrywanie wycieku promptów, eksfiltracji danych, nieautoryzowanych narzędzi i złośliwych dokumentów.
Nadzór człowieka	Możliwość wyłączenia modelu, wycofania dokumentu/promptu, zmiany wersji i oznaczenia odpowiedzi jako niebezpiecznej.
Monitoring jakości	Trafność, cytowania, odmowy, błędy, czas odpowiedzi i wykorzystanie zasobów bez ujawniania danych chronionych.
SBOM i licencje	Pełny wykaz modeli, bibliotek, komponentów, wersji, zależności i licencji.
Brak vendor lock-in	Prawo utrzymania i rozwoju przez podmiot trzeci; brak nieujawnionych zależności, aktywacji internetowej i utraty funkcji offline.
Zmiany produkcyjne	Zmiana modelu, checkpointu, precyzji, promptu systemowego, komponentu dokumentowego lub indeksu RAG wymaga regresji i zatwierdzenia.
Zgodność prawna	Ocena AI Act, RODO i prawa autorskiego; wymagane rejestry, informacje, ograniczenia i środki nadzoru.
Dostępność cyfrowa	WCAG 2.1 AA dla interfejsu i dokumentów użytkowych; audyt automatyczny i manualny, klawiatura, fokus, kontrast, semantyka, komunikaty błędów, alt text i czytnik ekranu.

Obszar	Wymagania do uwzględnienia
Kod źródłowy	Kod wszystkich komponentów wytworzonych lub zmodyfikowanych, historia repozytorium, tagi, build/deploy offline, konfiguracje, migracje, schematy API, prompty, guardrails, testy, lockfile, SBOM i dokumentacja developerska.
Prawa	Zasadą jest przeniesienie praw do elementów wytworzonych/zmodyfikowanych, wraz z prawami zależnymi i możliwością dalszego przekazania Ministrowi; licencja wyłączna tylko gdy przeniesienie praw jest prawnie niemożliwe. Komponenty standardowe pozostają na ujawnionych licencjach.
Oznaczenia finansowania	Nazwa i logo Ministra/MNiSW oraz informacja o finansowaniu w warstwie użytkowej, dokumentacji, materiałach szkoleniowych, raportach, szablonach eksportowych i publikacjach; bez surowych logów, telemetrii i API.

10. Testy i minimalne metryki odbiorowe

Zasada RFI: Na etapie RFI respondent nie wykonuje benchmarków ani UAT, lecz musi uwzględnić w cenie przygotowanie, wykonanie, remediację, regresję, raporty i wsparcie wszystkich testów opisanych poniżej.

Metryka	Minimalny poziom dla modelu
HLE text-only	>=35,0% dla modelu.
GPQA-Diamond	>=89,0% dla dokładnego checkpointu w docelowej kwantyzacji, pass@1, bez adaptacji i RAG.
MMLU-Pro albo AIME 2026	MMLU-Pro >=81,0% albo AIME 2026 >=93,0% bez narzędzi.
Retencja jakości	Średnia >=97,5%; spadek w każdym benchmarku <=3,0 p.p.
IAL - jakość Lean	>=85/100; żadna kategoria krytyczna <80%.
Ślepy test ekspertów	>=4,25/5; >=5 ekspertów, >=150 porównań A/B, >=3 oceny każdego porównania, łącznie >=450 ocen.
Poprawność cytowania	>=90%.
Pokrycie cytowaniem	>=95%.
Odmowa / brak podstaw	>=95%.
Halucynacje krytyczne	Brak potwierdzonych, nierozwiązanych przypadków w końcowym przebiegu podzbioru krytycznego po remediacji.
Poprawność JSON	>=98% odpowiedzi zgodnych ze schematem.
Recall@5 RAG	>=0,85.
Kontekst 256k	Trafność odnalezienia i cytowania >=90%; degradacja względem 32k <=5 p.p.
Dokumenty i grafika	>=90% pól krytycznych.
Wpływ adaptacji	>=+10 p.p. IAL względem bazowego modelu bez RAG albo zaakceptowany, uzasadniony brak regresji.
Testy bezpieczeństwa	>=95%; brak nierozwiązanych przypadków krytycznych w końcowym przebiegu.
Pamięć	>=384 GB nominalnie lub więcej według sizingu; rezerwa >=5% podczas S2.
Wywołania zewnętrzne AI	0.
Docelowa kwantyzacja	Co najmniej 4-bitowa; zgodność artefaktu, mapy precyzji i sum kontrolnych.
Stabilność	72 h; 0 restartów, OOM, zakleszczeń i błędów krytycznych; skuteczność żądań >=99,5%.

10.1. Struktura końcowego UAT - 300 scenariuszy

Obszar	Liczba	Zakres
Analiza procesu Lean	120	Struktura procesu i role 25; marnotrawstwo 20; wąskie gardła 20; 5 Why/Ishikawa 20; A3 15; VSM 10; warianty usprawnień 10.
RAG i cytowania	60	Wyszukiwanie właściwych fragmentów 25; poprawność cytowania 20; brak źródła, sprzeczności i wersjonowanie 15.
Język polski i dydaktyka	30	Poprawność, naturalność, terminologia Lean, czytelność, poziom szczegółowości i pytania naprowadzające.
JSON i narzędzia	25	Schematy, wywołania narzędzi, obsługa błędów i deterministyczne pola obowiązkowe.
Dokumenty i multimodalność	25	PDF tekstowy i skanowany, tabela, wykres, schemat, obraz strony i dokument wieloźródłowy.
Długi kontekst	15	Odnajdywanie, łączenie i cytowanie informacji w 128k oraz porównanie z 32k.
Bezpieczeństwo i zakres	25	Prompt injection, złośliwe dokumenty, wyciek danych, eskalacja uprawnień, pytania kliniczne, brak podstaw i próby obejścia.

Obszar	Liczba	Zakres
Łącznie	300	Każdy scenariusz: ID, kategoria, krytyczność, wejście, oczekiwane elementy, błędy niedopuszczalne, rubryka i dane dowodowe.

10.2. Zakres testów wliczony w cenę

- testy funkcjonalne: przypadki pozytywne, negatywne i wyjątkowe;
- ocena domenowa Lean oraz porównanie modelu bazowego, wariantu z RAG i wariantu po adaptacji z RAG;
- testy RAG, dokumentów i grafiki, długiego kontekstu, JSON i narzędzi;
- testy bezpieczeństwa, braku usług zewnętrznych i zakresu nieklinicznego;
- testy integracyjne: symulator, platforma szkoleniowa, repozytorium, adapter modelowy, monitoring/SIEM i - jeżeli zadeklarowane - SSO;
- odbiór infrastruktury, pamięci, storage, zasilania, backupu, licencji, sizingu i architektury;
- walidacja kwantyzacji, tożsamości modelu, manifestu i sum kontrolnych;
- test stabilności 72 h oraz konwersja środowiska testowego do produkcji;
- audyt WCAG 2.1 AA;
- remediacja niezgodności, testy regresji i raporty końcowe.

Zamawiający zamraża materiał UAT przed testem. Wykonawca dostarcza manifest profilu produkcyjnego, automatyzuje wykonanie, zapisuje odpowiedzi i logi bez ręcznej edycji oraz przygotowuje raporty surowe i zbiorcze.

11. Produkty i artefakty objęte wyceną

Kod	Produkt	Minimalny zakres
P.LEAN.01	Architektura rozwiązania	Architektura logiczna i techniczna, diagramy, przepływy danych, model bezpieczeństwa i sizing.
P.LEAN.02	Raport benchmarku modeli	Porównanie konfiguracji kandydackich: wyniki, precyzja wag, retencja, pamięć, koszt, wydajność, licencje i rekomendacja.
P.LEAN.03	Zbiór danych domenowych	Zestawy treningowe i walidacyjne oraz odrębne scenariusze UAT wraz z data card, pochodzeniem i statystykami.
P.LEAN.04	Model, model douczony/adapter domenowy	Wagi lub adapter, tokenizer, konfiguracja, model card i instrukcja uruchomienia.
P.LEAN.05	Baza wiedzy RAG	Repozytorium, indeksy, embeddingi, pipeline dokumentowy, panel publikacji i raport jakości retrieval.
P.LEAN.06	Biblioteka promptów	Minimum 25 walidowanych szablonów wraz z testami i historią wersji.
P.LEAN.07	Aplikacja i integracje	Interfejs, panel administracyjny, API, symulator, platforma szkoleniowa i pozostałe integracje.
P.LEAN.08	Raporty testów	Benchmarki i retencja oraz końcowy UAT: Lean, RAG, dokumenty, bezpieczeństwo, wydajność, integracje i zestawienie zarządcze.
P.LEAN.09	Dokumentacja	Instrukcje użytkownika, administratora, API, wdrożenia, backupu, aktualizacji, bezpieczeństwa i utrzymania.
P.LEAN.10	Szkolenia i przekazanie	Szkolenia, materiały i protokół przekazania do utrzymania.
P.LEAN.11	Kompletna infrastruktura on-premise	Sprzęt, pamięć, storage, sieć, montaż, zasilanie, monitoring, backup, licencje, instalacja i uruchomienie.
P.LEAN.12	Konwersja środowiska do produkcji	Oczyszczenie danych, rotacja sekretów, hardening, kopia bazowa, odtworzenie, izolacja i uruchomienie produkcyjne.
P.LEAN.13	Pakiet kodu źródłowego i utrzymaniowego	Repozytorium, historia, tag wydania, build/deploy offline, konfiguracja, migracje, testy, SBOM, licencje, dokumentacja developerska i prawa.

11.1. Minimalna dokumentacja

- architektura, przepływy danych, bezpieczeństwo, konfiguracja jednego środowiska i zależności;
- dokumentacja infrastruktury: BOM, numery seryjne, połączenia, miejsce instalacji, moc, chłodzenie, adresacja, firmware, licencje, gwarancje i schematy;
- obliczenie sizingowe: wagi, format, cache, bufor, kontekst, współbieżność, storage i rezerwa;

- model card, data card, rejestr ryzyk AI i karta ograniczeń;
- proces adaptacji/fine-tuningu lub adaptera, parametry, metryki i reprodukcja;
- pakiet benchmarkowy i manifesty, skrypty, logi, wyniki surowe i sumy kontrolne;
- opis RAG, parserów, OCR/VLM, metadanych, indeksacji, embeddingów, rerankingu i publikacji;
- instrukcje użytkownika, administratora, aktualizacji offline, backupu, odtworzenia, monitoringu i serwisu;
- repozytorium kodu i pełny pakiet utrzymaniowy przekazywany przy odbiorach rocznych i końcowym, w terminie umożliwiającym raportowanie do MNiSW.

11.2. Szkolenia

Grupa	Minimum	Zakres
Administratorzy AI i IT	8 godzin	Instalacja, konfiguracja, monitoring, bezpieczeństwo, backup, aktualizacja modeli i RAG.
Eksperci Lean i redaktorzy wiedzy	8 godzin	Walidacja źródeł, promptów i danych treningowych oraz udział w UAT Lean.
Trenerzy	4 godziny	Prowadzenie ćwiczeń, scenariusze, raporty i praca z symulatorem.
Użytkownicy końcowi	2 godziny	Bezpieczna praca z asystentem, formułowanie pytań, weryfikacja źródeł i ograniczenia.

12. Ramy realizacji, gwarancja i serwis

Szczegółowy harmonogram i płatności zostaną określone w przyszłym postępowaniu. Dla potrzeb RFI respondent podaje łączny czas realizacji, czas dostawy infrastruktury, czas do uruchomienia testowego i produkcyjnego oraz terminy krytyczne.

Okres	Zakres rezultatów
2026 - koncepcja i pilotaż	Architektura, wybór modelu, benchmarki, korpus, prototyp RAG/aplikacji i uruchomienie testowe.
2027 - wdrożenie	Adaptacja domenowa, aplikacja i integracje, UAT, test 72 h, odbiór i uruchomienie produkcyjne.
2028 - ewaluacja i stabilizacja	Ewaluacja, regresja, stabilizacja, aktualizacja wiedzy i raporty końcowe.

Obszar	Wymaganie
Gwarancja infrastruktury	Co najmniej 36 miesięcy od odbioru produkcyjnego.
Gwarancja oprogramowania	Co najmniej 12 miesięcy od odbioru produkcyjnego.
Zakres	Usuwanie błędów, poprawki bezpieczeństwa, aktualizacja dokumentacji, wsparcie odtworzenia i regresyjne potwierdzanie jakości po zmianach.
Cena	Pełny wymagany okres gwarancji i serwisu ma być zawarty w jednej cenie, również w części przypadającej po 31 grudnia 2028 r.
Aktualizacje	Podpisane pakiety offline w kontrolowanym repozytorium Zamawiającego.

Założenie SLA do jednolitej wyceny RFI: Finalny OPZ zawiera rozbieżne czasy w tabeli gwarancyjnej i Załączniku G.3. Aby odpowiedzi RFI były porównywalne, należy przyjąć wariant bardziej wymagający: błąd krytyczny - reakcja do 4 godzin roboczych i obejście/przywrócenie do 24 godzin; błąd istotny - reakcja do 1 dnia roboczego i naprawa do 5 dni roboczych; błąd pozostały/konsultacja - reakcja do 2 dni roboczych, termin naprawy uzgodniony. Ostateczna wartość zostanie jednoznacznie potwierdzona w SWZ.

13. Wymagany zakres odpowiedzi Respondenta

Odpowiedź może być syntetyczna, lecz musi pozwalać zweryfikować, że jedna cena obejmuje wszystkie istotne parametry niniejszego RFI. Należy przekazać:

- dane Respondenta i osoby kontaktowej;
- jednoznaczne TAK/NIE dla możliwości realizacji pełnego zakresu;
- opis jednej architektury w objętości maksymalnie 5 stron;
- dokładną nazwę modelu, checkpoint, licencję, rozmiar, parametry całkowite/aktywne, docelową kwantyzację i mapę precyzji;
- opis metody adaptacji: dostrojenie wag, PEFT/LoRA, adapter domenowy lub równoważne; plan danych i założenie dotyczące liczby przykładów;
- opis RAG: zakładany korpus, liczba dokumentów/stron/GB, minimum 100 000 fragmentów, embeddingi, reranker, OCR/VLM, storage i wydajność;
- opis infrastruktury, sizingu, pamięci, storage, parallelizmu, sieci, liczby urządzeń, mocy, chłodzenia i gwarancji;
- przewidywane parametry S1/S2, kontekstu, stabilności i dostępności;

- niewyceniony jednostkowo wykaz głównych elementów sprzętu IT;
- orientacyjny czas realizacji, czas dostawy sprzętu i zależności po stronie Zamawiającego;
- założenia, wyłączenia i ryzyka mogące zmienić cenę lub termin;
- cenę wyłącznie w dwóch kategoriach: Sprzęt IT oraz Pozostały zakres zamówienia, a także cenę łączną;
- dokładność oszacowania i okres ważności cen, preferowany minimum 90 dni.

Uproszczona cena: Nie należy podawać cen jednostkowych urządzeń, szczegółowego BOM cenowego, stawek osobodni, cen produktów P.LEAN, podziału ceny na lata, TCO ani osobnych cen opcji. Wymagane są tylko dwie kwoty składowe oraz cena łączna.

13.1. Zasady kalkulacji

- Cena w PLN netto, stawka i kwota VAT oraz wartość brutto.
- Sprzęt IT: fizyczne urządzenia i elementy infrastruktury przekazywane Zamawiającemu wraz z akcesoriami, połączeniami, montażem, zasilaniem i gwarancją producenta.
- Pozostały zakres: oprogramowanie, licencje, modele, analiza, architektura, instalacja, konfiguracja, adaptacja, RAG, aplikacja, API, integracje, testy, UAT, dokumentacja, szkolenia, prawa, gwarancja i serwis.
- Łączna cena obejmuje transport, ubezpieczenie dostawy, instalację, uruchomienie, odbiory i wszystkie świadczenia niezbędne do kompletnego wykonania.
- Każde wyłączenie lub założenie wpływające na cenę należy wskazać. Brak informacji oznacza, że cena obejmuje pełny zakres RFI.

14. Sposób i termin przekazania odpowiedzi

Pole	Informacja
Termin zadawania pytań	11.08.2026
Termin przekazania odpowiedzi	13.08.2026
Adres e-mail	adam.kaniuk@pimmswia.gov.pl
Tytuł wiadomości	RFI - Moduł AI Lean - [nazwa Respondenta]
Format	Wypełniony dokument edytowalny DOCX lub równoważny oraz PDF. Dopuszcza się schemat architektury i karty katalogowe.
Osoba kontaktowa	Adam Kaniuk

Materiały pomocnicze nie zastępują wypełnienia formularza cenowego i jednoznacznego potwierdzenia kompletności jednego wariantu.

Załącznik 1. Formularz odpowiedzi Respondenta

Instrukcja: Pola nieadekwatne należy oznaczyć „nie dotyczy” i uzasadnić. Odpowiedź dotyczy jednego kompletnego wariantu zgodnego z niniejszym RFI. Każde „NIE” lub odstępstwo należy opisać w części F.

A. Dane Respondenta

Pole	Odpowiedź
Pełna nazwa podmiotu	
Adres / siedziba	
NIP / REGON / nr rejestrowy	
Osoba kontaktowa	
E-mail / telefon	
Rola w potencjalnej realizacji	integrator / producent / dostawca infrastruktury / konsorcjant / inne
Strona internetowa	
Data sporządzenia odpowiedzi	

B. Karta techniczna jednego wariantu

Parametr	Odpowiedź Respondenta
Możliwość realizacji pełnego zakresu	TAK / NIE
Dokładna nazwa modelu i checkpointu	
Producent / właściciel wag / licencja	
Architektura i parametry	dense / MoE / inna; parametry całkowite i aktywne
Rozmiar checkpointu	GB
Docelowa kwantyzacja i mapa precyzji	
Silnik inferencyjny	
Parallelism / organizacja zasobów	TP / PP / EP / inne
Efektywny kontekst	tokeny; minimum 128 000
Natywny kontekst - W	tokeny; czy >=262 144?
Metoda adaptacji domenowej	wagi / PEFT / LoRA / adapter / inne
Liczba przykładów treningowych	minimum 3 000 albo uzasadnienie mniejszego zbioru
Zakres RAG	liczba dokumentów / stron / GB / fragmentów; minimum 100 000 fragmentów
Embeddingi / reranker / OCR-VLM	nazwy i lokalny sposób uruchomienia
Liczba i forma urządzeń	
Łączna nominalna pamięć inferencyjna	GB; minimum 384
Rezerwa pamięci w S2	%; minimum 5
Pamięć masowa	TB netto i rezerwa; minimum 30% po odbiorze
Przewidywane S1	TTFT / TPS / pełna odpowiedź
Przewidywane S2	TTFT / łączny TPS / pełna odpowiedź
Maksymalny pobór mocy	kW
Wymagane chłodzenie i miejsce	
Orientacyjny czas realizacji	miesiące
Orientacyjny czas dostawy sprzętu	tygodnie / miesiące
Dokładność oszacowania	+/- ...%
Okres ważności cen	dni

C. Opis proponowanego rozwiązania

Proszę opisać jedną architekturę w maksymalnie 5 stronach: komponenty, przepływ danych, model i adaptację, RAG, aplikację, integrację, bezpieczeństwo, monitoring, backup, pracę offline, testy i utrzymanie.

Opis Respondenta

Opis Respondenta**D. Główne elementy sprzętu IT - bez cen jednostkowych**

Tabela służy wyłącznie potwierdzeniu kompletności konfiguracji. Nie podawać cen pojedynczych urządzeń.

Lp.	Urządzenie / klasa elementu	Ilość	Kluczowe parametry i gwarancja
1			
2			
3			
4			
5			
6			
7			
8			

E. Założenia, wyłączenia, zależności i ryzyka

Należy wskazać wyłącznie elementy mające realny wpływ na wykonalność, cenę lub termin. Nie przedstawiać alternatywnych wariantów.

Lp.	Założenie / wyłączenie / ryzyko	Wpływ i sposób ograniczenia
1		
2		
3		
4		
5		
6		
7		

Załącznik 2. Formularz oszacowania ceny

Jedna wycena: Respondent podaje cenę jednego kompletnego wariantu. Koszt może być rozdzielony wyłącznie na Sprzęt IT i Pozostały zakres zamówienia. Nie należy dodawać innych pozycji cenowych.

- Sprzęt IT: fizyczne urządzenia i elementy infrastruktury przekazywane Zamawiającemu, w tym akcesoria, połączenia, montaż, zasilanie i gwarancja producenta.
- Pozostały zakres zamówienia: wszystkie pozostałe świadczenia i koszty, w tym oprogramowanie, licencje, modele, architektura, instalacja, adaptacja domenowa, RAG, aplikacja, API, integracje, testy, UAT, dokumentacja, szkolenia, prawa, gwarancja i serwis.

Lp.	Kategoria	Netto [PLN]	VAT [%]	VAT [PLN]	Brutto [PLN]
1	Sprzęt IT				
2	Pozostały zakres zamówienia				
	RAZEM				

A. Informacje o oszacowaniu

Pole	Odpowiedź
Charakter ceny	budżetowa / katalogowa / oszacowanie eksperckie / inna
Dokładność oszacowania	+/- ...%
Okres ważności ceny	... dni
Waluta	PLN
Czy cena obejmuje pełny zakres RFI?	TAK / NIE - w przypadku NIE opisać wyłączenie w Załączniku 1F
Czy cena obejmuje transport, instalację, uruchomienie i odbiory?	TAK / NIE
Czy cena obejmuje pełny wymagany okres gwarancji i serwisu?	TAK / NIE
Czy cena obejmuje przygotowanie i wykonanie wszystkich testów i UAT?	TAK / NIE

B. Uwagi do ceny

Wyjaśnienia dotyczące kalkulacji, podatków, kursów walut lub istotnych warunków cenowych

Załącznik 3. Oświadczenie Respondenta

- ☐ Przekazane informacje mają charakter orientacyjny i zostały przygotowane z należytą starannością na podstawie niniejszego RFI.
- ☐ Odpowiedź dotyczy jednego kompletnego wariantu rozwiązania.
- ☐ W cenie uwzględniono wszystkie elementy niezbędne do uruchomienia, odbioru i eksploatacji, z wyjątkiem wyraźnie opisanych wyłączeń.
- ☐ Potwierdzono wszystkie istotne parametry modelu, adaptacji, RAG, infrastruktury, testów, dokumentacji, praw, gwarancji i serwisu.
- ☐ Wszystkie założenia, zależności, wyłączenia i ryzyka wpływające na cenę lub termin zostały wskazane.
- ☐ Respondent wyraża zgodę na kontakt w celu wyjaśnienia odpowiedzi i wykorzystanie informacji do przygotowania postępowania, z zastrzeżeniem prawidłowo oznaczonej tajemnicy przedsiębiorstwa.
- ☐ Odpowiedź nie stanowi oferty handlowej ani zobowiązania do zawarcia umowy.

Pole	Odpowiedź
Imię i nazwisko	
Stanowisko	
Data	
Podpis / podpis elektroniczny	

KONIEC FORMULARZA RFI