

WYKONAWCY

Dot. zapytania ofertowego w zakresie zamówienia pn.: „Dostawa i wdrożenie narzędzia do transkrypcji mowy do tekstu”.

Zamawiający przekazuje treść pytań oraz wyjaśnienia i dokonane przez Zamawiającego zmiany treści zapytania ofertowego.

Pytanie nr 1

Zgodność Whisper on-premise z wymaganiem pkt III ust. 1 OPU

Zamawiający zmienił wymaganie z pkt III ust. 1 OPU, wskazując w pkt 1 zaktualizowanego zapisu, że Narzędzie „nie wykonuje w trakcie pracy operacji opartych na sieciach neuronowych ani modelach uczenia maszynowego (inferencja) w środowisku produkcyjnym”, a jednocześnie potwierdzając w odpowiedziach na pytania nr 4 i 5, że:

„wymóg odnosi się do sposobu działania Narzędzia w środowisku produkcyjnym Zamawiającego”, oraz że „spełnienie powyższych warunków przez Narzędzie działające wyłącznie lokalnie w infrastrukturze Zamawiającego jest wystarczające do uznania zgodności z wymaganiami sekcji III Załącznika nr 1”.

Zamierzamy zaoferować Narzędzie oparte na modelu Whisper (OpenAI, licencja MIT), wdrożonym w całości on-premise na infrastrukturze UKNF, spełniającym łącznie następujące warunki:

1. Model akustyczny wytworzony przez producenta (OpenAI) poza środowiskiem Zamawiającego, przed dostawą — zgodnie z pkt III ust. 1 ppkt 5 OPU;
2. W trakcie pracy Narzędzia w środowisku produkcyjnym Zamawiającego wykonywana jest inferencja (wnioskowanie przez zamrożony model transformer) — tj. przetwarzanie nagrania do tekstu jest technicznie realizowane przez sieć neuronową działającą lokalnie na GPU Zamawiającego;
3. Narzędzie nie uczy się na danych Zamawiającego (brak fine-tuningu, brak uczenia online);
4. Narzędzie nie wysyła żadnych danych poza infrastrukturę Zamawiającego;
5. Wyniki są deterministyczne i audytowalne (decoding greedy/beam search, temperature=0).

Dostzegamy napięcie pomiędzy literalnym brzmieniem pkt III ust. 1 ppkt 1 OPU (zakaz inferencji w środowisku produkcyjnym) a intencją potwierdzoną w odpowiedziach na pytania nr 4 i 5, gdzie Zamawiający wskazał, że celem jest wyłącznie wykluczenie narzędzi wysyłających dane do chmury, uczących się na danych Zamawiającego lub generujących wyniki niedeterministyczne.

Zwracamy się z prośbą o jednoznaczne potwierdzenie, czy Zamawiający zaakceptuje Narzędzie oparte na modelu Whisper on-premise, wykonujące lokalną inferencję na GPU

Zamawiającego, spełniające warunki z pkt III ust. 1 ppkt 2–5 OPU, w świetle intencji określonej w odpowiedziach na pytania nr 4 i 5.

Odpowiedź

Zamawiający zmienia treść pkt 1 w ust. 1 w części III Załącznika nr 1 do umowy w następujący sposób:

„1) Narzędzie musi zapewnić wykonywanie inferencji sieci neuronowych w pełni lokalnie, w środowisku produkcyjnym Zamawiającego bez zależności od zewnętrznych usług w tym usług chmurowych oraz bez przesyłania danych poza infrastrukturę Zamawiającego.”.

Pytanie nr 2

Wsparcie producenta dla oprogramowania open-source (MIT)

Zgodnie z §4 ust. 8 oraz §5 ust. 2–3 Umowy, Wykonawca zobowiązany jest do zapewnienia wsparcia producenta, w tym dostarczenia dokumentu wystawionego przez producenta Narzędzia poświadczającego zapewnienie licencji oraz wsparcia na okres 12 miesięcy, wraz z dostępem do portalu producenta.

Oferowane Narzędzie będzie oparte na oprogramowaniu open-source wydanym na licencji MIT (m.in. Whisper, pyannote.audio). W przypadku takiego oprogramowania:

- producent nie wystawia formalnych dokumentów licencyjnych w tradycyjnym rozumieniu — licencja jest udzielana automatycznie na podstawie pliku LICENSE w repozytorium;
- nie istnieje portal producenta z panelem zarządzania licencjami;
- wsparcie producenta ma charakter publiczny (GitHub Issues, releases, changelog).

Prosimy o potwierdzenie, że w przypadku komponentów open-source (licencja MIT) za wystarczające spełnienie wymogu §5 ust. 2–3 Umowy Zamawiający uzna:

- a) dostarczenie kopii pliku LICENSE z repozytorium producenta (OpenAI/Whisper);
- b) dostarczenie linku/wydruku PDF do publicznego repozytorium GitHub producenta z aktualną wersją oprogramowania i historią wydań (changelog);
- c) pisemne oświadczenie Wykonawcy o zapewnieniu aktualizacji i wsparcia technicznego w ramach serwisu przez cały okres Umowy.

W nawiązaniu do opublikowanego projektu umowy (Załącznik nr 2 do zapytania ofertowego) oraz dokumentu wyjaśnień z dnia 30 kwietnia 2026 r., zwracamy się z prośbą o doprecyzowanie poniższych zapisów umownych, które — w obecnym brzmieniu — uniemożliwiają rzetelną wycenę ryzyk i przygotowanie odpowiedzialnej oferty.

Odpowiedź

Zamawiający zmienia projekt umowy w zakresie, o którym mowa w pytaniu. Projekt umowy po zmianie stanowi Załącznik nr 1 do pisma.

Ponadto Zamawiający zmienia treść wzoru formularza ofertowego poprzez dodanie oświadczenia do wypełnienia przez Wykonawcę o następującej treści:

„Oferuję rozwiązanie oparte o licencje (należy wyszczególnić wszystkie licencje, o które oparte jest oferowane rozwiązanie.*

* W przypadku jeśli Wykonawca oferuje oba typy licencji, tj. oprogramowanie oparte o licencje typu open source oraz o licencje wieczyste to należy również wskazać, który typ licencji jakiego zakresu zamówienia dotyczy.”

Zmodyfikowany formularz ofertowy stanowi Załącznik nr 2 do pisma.

Pytanie nr 3

Termin 30 dni na wdrożenie a faktyczna sekwencja obowiązków (§2 ust. 2 w zw. z §4 ust. 1 i §6)

Zgodnie z §2 ust. 2 Wykonawca zobowiązany jest dostarczyć i wdrożyć Narzędzie w terminie 30 dni od udostępnienia środowiska przez Zamawiającego.

Jednocześnie projekt umowy przewiduje następującą sekwencję zdarzeń poprzedzających wdrożenie, która musi się zmieścić w tych samych 30 dniach:

- §4 ust. 1: Wykonawca w ciągu 21 dni roboczych od otrzymania szablonów opracowuje Projekty;
- §4 ust. 2: Zamawiający ma 7 dni roboczych na akceptację Projektów (z możliwością 3 iteracji);
- §6 ust. 5: Zamawiający przeprowadza testy bezpieczeństwa w ciągu 18 dni roboczych;
- §6 ust. 15: Zamawiający przeprowadza testy akceptacyjne w ciągu 10 dni roboczych.

Samo zatem minimum proceduralnej obsługi po stronie Zamawiającego wynosi 35 dni roboczych (~7 tygodni kalendarzowych), nie licząc czasu pracy Wykonawcy, co czyni termin 30 dni kalendarzowych technicznie niemożliwym do dotrzymania bez winy Wykonawcy.

Prosimy o potwierdzenie, że:

1. Termin 30 dni z § 2 ust. 2 jest liczony od momentu udostępnienia środowiska gotowego zgodnie z zaakceptowanym projektem technicznym i nie obejmuje czasu przeznaczonego na prace po stronie Zamawiającego (analogicznie do zasady wyrażonej w §6 ust. 13 i 19);
2. Zamawiający rozważy wydłużenie terminu wdrożenia do 60 dni roboczych od udostępnienia środowiska, uwzględniając realną sekwencję procesów projektowych, testów bezpieczeństwa i akceptacyjnych;
3. Bieg terminu ulega zawieszeniu na czas oczekiwania na decyzje/akceptacje Zamawiającego w każdym etapie przewidzianym w § 4 i § 6 Umowy.

Odpowiedź

Ad 1. Zamawiający wyjaśnia, że zgodnie z treścią § 2 ust. 2 projektu umowy termin 30 dni roboczych liczony jest od udostępnienia środowiska przez Zamawiającemu, po akceptacji projektu technicznego, a co za tym idzie termin 21 dni roboczych, o którym mowa w § 4 ust. 1 projektu umowy, nie wlicza się do terminu, o którym mowa w § 2 ust. 2 projektu umowy;

Ad 2. Zamawiający nie uwzględnia wniosku;

Ad 3. Zamawiający wyjaśnia, że terminy realizacji umowy wskazane w § 4 projektu umowy są obowiązujące dla Stron. Zamawiający określił w § 4 projektu umowy czasy realizacji umowy po stronie Wykonawcy jak i po stronie Zamawiającego. W przypadku § 6 projektu umowy dotyczącego testów bezpieczeństwa i testów akceptacyjnych, ust. 13 oraz ust. 19 wskazuje, że czas przeznaczony na wykonanie wszelkich prac po stronie Zamawiającego nie jest wliczany do terminu określonego w § 2 ust. 2 projektu umowy.

Pytanie nr 4

Harmonogram i etapowość płatności (§8)

Projekt umowy przewiduje jednorazową płatność za poz. 1 (dostawa, wdrożenie, licencja, wsparcie producenta) wyłącznie po podpisaniu protokołu odbioru końcowego (§8 ust. 4). Oznacza to, że Wykonawca przez cały okres wdrożenia (który — jak wskazano w pytaniu A — może realnie trwać kilka miesięcy) ponosi pełne koszty realizacji bez jakiegokolwiek płatności przejściowej.

Stanowi to istotne ryzyko finansowe dla Wykonawcy, nieproporcjonalne do wartości zamówienia poniżej progów ustawy Pzp.

Prosimy o rozważenie i potwierdzenie możliwości wprowadzenia etapowego harmonogramu płatności za poz. 1, np.:

- 50% wynagrodzenia — po podpisaniu protokołu potwierdzającego pozytywny wynik testów bezpieczeństwa (§6 ust. 8 Umowy);
- 50% wynagrodzenia — po podpisaniu protokołu odbioru końcowego (§5 ust. 1 Umowy).

Ewentualnie prosimy o wskazanie, czy Zamawiający dopuszcza inne, zaproponowane przez Wykonawcę rozwiązania w zakresie płatności etapowych.

Odpowiedź

Zamawiający nie uwzględnia wniosku.

Pytanie nr 5

Zawieszenie biegu terminów przy niezapewnieniu dostępu przez Zamawiającego (§3 ust. 6 pkt 5 i §9)

§3 ust. 6 pkt 5 zobowiązuje Wykonawcę do realizacji prac w siedzibie Zamawiającego pod jego nadzorem, z 10-dniowym wyprzedzeniem przy zawiadomieniu o planowanych pracach.

§9 ust. 8 wymaga z kolei pisemnej zgody Zamawiającego na każdy przypadek pracy zdalnej.

W przypadku gdy Zamawiający:

- nie zapewni w uzgodnionym terminie dostępu do siedziby, lub
- nie wyrazi zgody na pracę zdalną w sytuacji, gdy dostęp fizyczny jest niemożliwy (np. dni wolne, awaria, procedury bezpieczeństwa KNF),

— Wykonawca może ponosić kary z tytułu opóźnienia, do którego sam nie doprowadził.

Prosimy o potwierdzenie, że:

1. W przypadku opóźnienia po stronie Zamawiającego w zapewnieniu dostępu do siedziby lub środowiska, bieg terminów realizacji ulega automatycznemu zawieszeniu o czas tego opóźnienia;
2. Zamawiający wyraża generalną zgodę na pracę zdalną w zakresie prac niewymagających fizycznego dostępu do infrastruktury KNF (tj. prace programistyczne, dokumentacyjne, testy jednostkowe), bez konieczności uzyskiwania odrębnej zgody każdorazowo.

Odpowiedź

Ad. 1 Zamawiający zwraca uwagę, że Wykonawca nie wskazał konkretnie biegu jakich terminów dotyczy pytanie. Zamawiający wyjaśnia, że termin określony dla projektu technicznego jest niezależny od zapewnienia dostępu przez Zamawiającego a termin dostawy i wdrożenia zależy od udostępnienia środowiska przez Zamawiającego, więc brak zapewnienia dostępu nie rozpoczyna tego terminu.

Ponadto Zamawiający nie przewiduje opóźnień w zapewnieniu dostępu dla Wykonawcy do siedziby Zamawiającego. Zamawiający nie ma interesu w opóźnianiu zapewnienia dostępu dla Wykonawcy, o którym mowa w pytaniu, i będzie zapewniał ten dostęp niezwłocznie zgodnie ze swoimi faktycznymi możliwościami. Zamawiający nie przewiduje braku możliwości dostępu fizycznego dla Wykonawcy w dni robocze.

Mając powyższe na uwadze Wykonawca powinien zaplanować prace w taki sposób aby prace wymagające dostępu do infrastruktury nie wymagały ich realizacji w dni wolne od pracy.

Ad. 2 Zgodnie z treścią § 9 ust. 8 projektu umowy zgodę na pracę zdalną Zamawiający udziela w uzasadnionych przypadkach, uniemożliwiających lub utrudniających realizację umowy na dotychczasowych zasadach, na uzasadniony wniosek Wykonawcy.

Pytanie nr 6

Limit odpowiedzialności Wykonawcy (§12)

§12 ust. 15 przewiduje możliwość kumulatywnego naliczania kar za wiele tytułów jednocześnie, bez wyłączenia podwójnego karania za to samo zdarzenie. §12 ust. 16 określa łączny limit kar na 50% wynagrodzenia (z czego za zwłokę — 30%).

Przy jednorazowej płatności (brak zaliczki) i łącznym wynagrodzeniu ok. 170 tys. zł netto, maksymalna ekspozycja na kary umowne wynosi ok. 85–95 tys. zł, co przy jednoczesnym braku jakichkolwiek zaliczek stanowi ryzyko nieproporcjonalne do skali projektu.

Ponadto §12 ust. 11 zachowuje prawo Zamawiającego do odszkodowania uzupełniającego na zasadach ogólnych ponad kwotę kar umownych — bez limitu kwotowego.

Prosimy o:

1. Potwierdzenie, że kary za zwłokę nie mogą być naliczane jednocześnie z karą za odstąpienie z tego samego tytułu (zasada non bis in idem);

2. Rozważenie wprowadzenia górnego limitu łącznej odpowiedzialności Wykonawcy (kary umowne + odszkodowanie uzupełniające) na poziomie 100% wynagrodzenia netto z poz. 1 Formularza ofertowego;
3. Potwierdzenie, że kara za zwłokę (§12 ust. 3 — 500 zł/dzień) jest jedyną karą za nieterminowe wdrożenie i nie kumuluje się z innymi karami za ten sam okres.

Odpowiedź

Ad 1. Zamawiający wyjaśnia, że w przypadku, gdy odstąpienie od Umowy nastąpiłoby z przyczyn leżących po stronie Wykonawcy, stanowiących jednocześnie podstawę do naliczenia kary umownej za zwłokę, kary te nie będą kumulowane. W takim przypadku zostanie naliczona kara umowna z tytułu odstąpienia od Umowy z przyczyn leżących po stronie Wykonawcy. Powyższe nie wyłącza możliwości naliczenia kar umownych za zwłokę w przypadku innych naruszeń obowiązków umownych, które nie stanowią podstawy do odstąpienia od Umowy.

Ad 2. Zamawiający nie uwzględnia wniosku.

Ad 3. Zamawiający wyjaśnia, że kara umowna określona w § 12 ust. 3 projektu umowy dotyczy zwłoki w dostarczeniu i wdrożeniu Narzędzia, co wynika wprost z treści tego postanowienia. Naliczenie tej kary nie wyłącza możliwości naliczenia innych kar umownych, w tym za zwłokę, przewidzianych w § 12, w przypadku wystąpienia innych naruszeń obowiązków umownych w tym samym okresie.

Pytanie nr 7

Aktualizacje komponentów open-source a zakres serwisu (§4 ust. 9 i §6 ust. 2 pkt 17)

§6 ust. 2 pkt 17 zobowiązuje Wykonawcę do utrzymywania Narzędzia wolnego od znanych podatności CVE przez cały okres wsparcia. Jednocześnie §4 ust. 9 ogranicza serwis do 20 roboczogodzin miesięcznie bez możliwości przenoszenia na kolejny miesiąc.

W praktyce przy stosie technologicznym opartym na komponentach open-source (Python, biblioteki ML, React), nowe podatności CVE publikowane są kilka razy w miesiącu. Każda aktualizacja biblioteki bezpieczeństwa może wymagać retestów — co przy limicie 20 rg/msc może okazać się niewystarczające, a Wykonawca nie będzie miał możliwości wykonania pracy niezbędnej do spełnienia wymogu z §6 ust. 2 pkt 17.

Prosimy o potwierdzenie, że:

1. Obowiązek usuwania podatności CVE w ramach wsparcia producenta (§4 ust. 8) jest odrębny od serwisu (§4 ust. 9) i nie wlicza się do limitu 20 rg/msc;
2. Zamawiający dopuszcza — w przypadku kumulacji zleceń wynikających z CVE — przeniesienie niewykorzystanych godzin serwisowych na kolejny miesiąc, lub rozliczenie nadwyżki w ramach modyfikacji (poz. 2 Formularza ofertowego);
3. Wymaganie z §6 ust. 2 pkt 17 dotyczy wyłącznie znanych podatności bezpieczeństwa dla komponentów i bibliotek bezpośrednio dostarczonych przez Wykonawcę, a nie warstwy systemowej RHEL pozostającej pod kontrolą Zamawiającego.

Odpowiedź

Zamawiający zmienia projekt umowy w zakresie, o którym mowa w pytaniu. Projekt umowy po zmianie stanowi Załącznik nr 1 do pisma.

Pytanie nr 8

Dotyczy: Wyjaśnienia i zmiany zapytania ofertowego z dnia 30 kwietnia 2026 r. – odpowiedź na Pytanie nr 1

Zamawiający zmienił ust. 1 w części III Załącznika nr 1 do projektu umowy (Opis Przedmiotu Umowy) wprowadzając m.in. następujące wymaganie:

„1. Narzędzie: 1) nie wykonuje w trakcie pracy operacji opartych na sieciach neuronowych ani modelach uczenia maszynowego (inferencja w środowisku produkcyjnym);”

Zgodnie z definicją IBM (<https://www.ibm.com/think/topics/ai-inference>), inferencja AI (ang. AI inference) to proces, w którym wytrenowany model stosuje wyuczone wzorce do generowania wyników dla nowych danych wejściowych. Innymi słowy – każde działanie wytrenowanego modelu w aplikacji produkcyjnej stanowi inferencję.

Wszelkie dostępne na rynku narzędzia do automatycznej transkrypcji mowy na tekst oraz diaryzacji (podziału wypowiedzi na mówców), w tym powszechnie stosowane rozwiązania takie jak Whisper czy Pyannote, opierają swoje działanie na modelach uczenia maszynowego i sieciach neuronowych, a zatem ich działanie w środowisku produkcyjnym nierozzerwalnie wiąże się z wykonywaniem inferencji. Modele te są kompilowane na podstawie skończonych zbiorów danych treningowych; dane przetwarzane przez użytkowników służą wyłącznie do wygenerowania żadanego wyniku i w żaden sposób nie są wykorzystywane do dalszego uczenia modelu.

Wprowadzone przez Zamawiającego wymaganie wyklucza zatem praktycznie wszystkie nowoczesne narzędzia transkrypcji dostępne na rynku i ogranicza możliwości realizacji zamówienia wyłącznie do rozwiązań opartych na przestarzałych technologiach sprzed ery uczenia maszynowego, charakteryzujących się znacząco niższą jakością i trafnością transkrypcji.

W związku z powyższym, wnosimy o wykreślenie wymagania zawartego w pkt III ust. 1 pkt 1 OPZ lub jego zastąpienie zapisem dopuszczającym stosowanie modeli AI wykonujących inferencję wyłącznie lokalnie (on-premise), bez dostępu do zewnętrznych serwerów lub usług chmurowych – co pozostaje w pełnej zgodności z wymaganiem pkt III ust. 1 pkt 3 OPZ.

Odpowiedź

Vide odpowiedź na pytanie nr 1.

Pytanie nr 9

Dotyczy: Wyjaśnienia i zmiany zapytania ofertowego z dnia 30 kwietnia 2026 r. – odpowiedź na Pytanie nr 1

Zamawiający zmienił ust. 1 w części III Załącznika nr 1 do projektu umowy (Opis Przedmiotu Umowy) wprowadzając m.in. następujące wymaganie:

„1. Narzędzie: 4) generuje wyniki w sposób deterministyczny i audytowalny;”

Wymaganie deterministyczności, rozumiane jako generowanie identycznego wyniku przy każdym uruchomieniu dla tych samych danych wejściowych, jest technicznie nieosiągalne dla narzędzi transkrypcji opartych na modelach probabilistycznych, jakimi są Whisper oraz Pyannote. Narzędzia te generują transkrypcję na podstawie rozkładów prawdopodobieństwa i nawet przy ustawieniu temperatury modelu równej zero (tj. wyłączeniu losowości) pełna deterministyczność na poziomie surowego wyjścia modelu nie jest gwarantowana przez producentów.

Jednocześnie narzędzie do transkrypcji nie jest systemem autonomicznym. Jego rolą jest wspomaganie pracy operatora, który weryfikuje wygenerowaną transkrypcję i wprowadza niezbędne korekty (m.in. w zakresie pisowni imion i nazwisk, nazw własnych, skrótów). To właśnie efekt końcowy procesu – transkrypcja po przeglądzie i korekcie operatora – stanowi właściwy, powtarzalny i audytowalny wynik pracy systemu. Trafność transkrypcji uzyskiwanej przez tego typu narzędzia wynosi do 98% (w zależności od jakości nagrania i języka), co czyni pracę operatora korektą o ograniczonym zakresie, a nie tworzeniem dokumentu od podstaw. Audytowalność systemu jest zapewniona poprzez przechowywanie w bazie danych surowej transkrypcji wygenerowanej przez model oraz pełnej historii modyfikacji wprowadzanych przez operatora aż do chwili wygenerowania raportu końcowego.

W związku z powyższym, wnosimy o doprecyzowanie, że:

- wymaganie deterministyczności dotyczy wyniku końcowego procesu transkrypcji, tj. efektu pracy operatora systemu, a nie surowego wyjścia generowanego przez model AI;
- wymaganie audytowalności jest spełnione poprzez przechowywanie w bazie danych zarówno surowej transkrypcji, jak i pełnej historii modyfikacji wprowadzanych przez użytkownika.

Odpowiedź

Zamawiający wyjaśnia, że wymaganie dotyczące generowania przez Narzędzie wyników w sposób deterministyczny oznacza, że Narzędzie zapewni powtarzalność procesu transkrypcji

poprzez zastosowania konfiguracji modelu AI minimalizującej losowość, gwarantując, że wynik końcowy procesu transkrypcji poddany korekcie operatora będzie merytorycznie spójny niezależnie od probabilistycznego charakteru modelu AI zastosowanego w Narzędziu.

Zamawiający wyjaśnia, że wymaganie audytowalności Narzędzia jest spełnione kiedy Narzędzie zapewnia przechowywanie w bazie danych zarówno danych wejściowych, surowej transkrypcji, jak i pełnej historii modyfikacji transkrypcji wprowadzanych przez użytkownika oraz metadanych technicznych dotyczących procesu wykonywania transkrypcji.

Pytanie nr 10

Czy Zamawiający dopuszcza zastosowanie narzędzia do transkrypcji, które działa w pełni lokalnie (on-premise), bez jakiegokolwiek komunikacji z zewnętrznymi usługami, wykorzystuje wcześniej wytrenowane modele (np. akustyczne i językowe), a jego działanie w środowisku Zamawiającego polega wyłącznie na ich wykonywaniu (inferencji) bez jakiegokolwiek uczenia, dostrajania ani wysyłania danych poza infrastrukturę?

W szczególności prosimy o potwierdzenie czy wykorzystanie udostępnionego przez Zamawiającego GPU (NVidia Tesla V100) do przyspieszenia takiego przetwarzania jest zgodne z wymaganiami OPZ.

Odpowiedź

Zmawiający wyjaśnia, że nie stawia wymagań w OPU dotyczących możliwości i zakresu wykorzystania udostępnionego przez Zamawiającego GPU (NVidia Tesla V100) do przyspieszenia takiego przetwarzania.

W pozostałym zakresie vide odpowiedź na pytanie nr 1.

Pytanie nr 11

Zamawiający zmienił wymaganie z pkt III ust. 1 OPU, wskazując w pkt 1 zaktualizowanego zapisu, że Narzędzie „nie wykonuje w trakcie pracy operacji opartych na sieciach neuronowych ani modelach uczenia maszynowego (inferencja) w środowisku produkcyjnym”, a jednocześnie potwierdzając w odpowiedziach na pytania nr 4 i 5, że:

„wymóg odnosi się do sposobu działania Narzędzia w środowisku produkcyjnym Zamawiającego”, oraz że „spełnienie powyższych warunków przez Narzędzie działające wyłącznie lokalnie w infrastrukturze Zamawiającego jest wystarczające do uznania zgodności z wymaganiami sekcji III Załącznika nr 1”.

Zamierzamy zaoferować Narzędzie oparte na modelu Whisper (OpenAI, licencja MIT), wdrożonym w całości on-premise na infrastrukturze UKNF, spełniającym łącznie następujące warunki:

1. Model akustyczny wytworzony przez producenta (OpenAI) poza środowiskiem Zamawiającego, przed dostawą — zgodnie z pkt III ust. 1 ppkt 5 OPU;
2. W trakcie pracy Narzędzia w środowisku produkcyjnym Zamawiającego wykonywana jest inferencja (wnioskowanie przez zamrożony model transformer) — tj. przetwarzanie nagrania do tekstu jest technicznie realizowane przez sieć neuronową działającą lokalnie na GPU Zamawiającego;
3. Narzędzie nie uczy się na danych Zamawiającego (brak fine-tuningu, brak uczenia online);
4. Narzędzie nie wysyła żadnych danych poza infrastrukturę Zamawiającego;
5. Wyniki są deterministyczne i audytowalne (decoding greedy/beam search, temperature=0).

Dostrzegamy napięcie pomiędzy literalnym brzmieniem pkt III ust. 1 ppkt 1 OPU (zakaz inferencji w środowisku produkcyjnym) a intencją potwierdzoną w odpowiedziach na pytania nr 4 i 5, gdzie Zamawiający wskazał, że celem jest wyłącznie wykluczenie narzędzi wysyłających dane do chmury, uczących się na danych Zamawiającego lub generujących wyniki niedeterministyczne.

Zwracamy się z prośbą o jednoznaczne potwierdzenie, czy Zamawiający zaakceptuje Narzędzie oparte na modelu Whisper on-premise, wykonujące lokalną inferencję na GPU Zamawiającego, spełniające warunki z pkt III ust. 1 ppkt 2–5 OPU, w świetle intencji określonej w odpowiedziach na pytania nr 4 i 5.

Odpowiedź

Vide odpowiedź na pytanie nr 1.

Ponadto Zamawiający wydłużył termin składania ofert. Ofertę należy złożyć w terminie do **5 czerwca 2026 r.**

Sławomir Flis

Dyrektor

Departamentu Informatyki

/podpisano kwalifikowanym podpisem elektronicznym/

Załączniki:

Załącznik nr 1 – zmodyfikowany projekt umowy;

Załącznik nr 2 – aktualny formularz ofertowy.