

OPIS PRZEDMIOTU ZAMÓWIENIA

Przedmiotem niniejszego zamówienia jest przeprowadzenie prac badawczo rozwojowych w ramach projektu pn. "Zintegrowany system analizy danych OSINT, identyfikacji ryzyk i oceny wiarygodności osób i podmiotów gospodarczych" (zwany dalej Integratorem).

Integrator służy do analizy danych udostępnionych z publicznie dostępnych źródeł na potrzeby ustalenia scoringu i oceny ryzyka badanego podmiotu oraz wykrywania anomalii w prowadzonej przez niego działalności. Analizuje dane jawnoźródłowe, które w zagregowany sposób, przy wykorzystaniu ogólnodostępnych rejestrów rządowych i komercyjnych pozwolą na analizę ryzyka badanego podmiotu. Zadaniem integratora danych jest połączenie praktycznie nieograniczonej liczby źródeł wiedzy o dowolnym charakterze i strukturze w celu przeprowadzenia analizy ryzyka/scoringu dla badanego podmiotu. Bazując na danych publicznie dostępnych w sieci Internet Integrator dokonuje identyfikacji sieci powiązań podmiotów i wynikających z tego ryzyk wraz z określeniem scoringu dla badanego podmiotu. Kluczowym elementem jest połączenie ogólnodostępnych rejestrów, baz danych, wykazów, danych statystycznych, raportów, które agregują informacje jawnoźródłowe. Integrator danych przy wykorzystaniu m.in. „fuzzy matchingu” pozwala na zbudowanie przybliżonych powiązań, relacji między obiektami, nawet w przypadku braku oczywistej zależności między nimi, wynikającej np. z różnego rodzaju zapisu danych. Cechą indywidualną integratora jest umożliwienie użytkownikowi dowolnej parametryzacji przy wykorzystaniu podpiętych źródeł wiedzy tzn., że użytkownik sam może podjąć decyzję jakie źródła powinny zostać przeanalizowane w danej sprawie dla sprawniejszego realizowania procesu Business Intelligence. Innowacyjność rozwiązania opiera się na zautomatyzowaniu procesu przygotowywania danych oraz gromadzeniu i konsolidacji informacji z wielu źródeł w czasie rzeczywistym.

Integrator umożliwia dostarczenie użytkownikowi bieżących raportów o zagrożeniach w (potencjalnej) współpracy z badanym podmiotem, prawdopodobieństwie ich wystąpienia oraz ewentualnych skutkach i rozmiarach ewentualnych strat. W oparciu o przedstawione wyniki w raporcie końcowym, użytkownik otrzyma informacje wskazujące na potencjalne ryzyka wynikające z prowadzonej działalności gospodarczej, będącej przedmiotem analizy.

Integrator wykorzystywać będzie elementy przetwarzania języka naturalnego (NLP), sztucznej inteligencji (AI) oraz uczenia maszynowego (ML) do zautomatyzowanego wykrywania anomalii dla danego podmiotu.

Integrator będzie realizował korelację i przetwarzanie następujących kategorii danych:

- dane pochodzące z wybranych ogólnodostępnych rejestrów rządowych i komercyjnych (np. KRS, CEIDG, Regon, CRBR, itp.), krajowych i zagranicznych
- wzorzec zachowań stale wzbogacany o dane rzeczywiste badanego podmiotu,
- dane wywiadowcze nieustrukturyzowane zbierane w oparciu o metody OSINT (Open Source Intelligence) w odniesieniu do badanego podmiotu,



Załącznik numer 3 do postępowania

- dane opisujące wzorzec zachowań podmiotów o podobnej specyfice do specyfiki badanego podmiotu (tzw. wzorzec branżowy, wertykalny),
- dane wywiadowcze nieustrukturyzowane zbierane w oparciu o metody OSINT w odniesieniu do podmiotów o podobnej specyfice do specyfiki badanego podmiotu,
- dodatkowe dane ustrukturyzowane lub nieustrukturyzowane zbierane w odniesieniu do badanego podmiotu.

Zakres prac do wykonania przez podwykonawcę (oferenta) w poszczególnych zadaniach w ramach projektu:

Etap 1 – Zadanie A – opracowanie koncepcji do analizy danych udostępnionych z różnych źródeł (zwanej dalej schematem danych)

W wyniku realizacji zadania powstanie szczegółowy opis struktury danych, który określi sposób ich gromadzenia na potrzeby ich przetwarzania przez Integrator. Dodatkowo schemat danych będzie określał uzgodnione ze Zleceniodawcą mechanizmy ich anonimizacji, parametry retencji danych, atrybuty wymagalności i schematy relacji. Schemat danych powinien uwzględniać również rekomendowane metody służące do zapewnienia poufności i integralności danych. Ponadto schemat danych powinien określać istotne wymagania względem środowiska informatycznego, na którym dane będą przetwarzane w ramach Zadania B w etapie 2.

W wyniku realizacji zadania pozyskiwane będą:

- dane z ogólnodostępnych źródeł rządowych i komercyjnych, m.in.: Krajowy Rejestr Sądowy, Centralna Ewidencja i Informacja o Działalności Gospodarczej, Centralny Rejestr Beneficjentów Rzeczywistych, Krajowy rejestr urzędowy podmiotów gospodarki narodowej (REGON), SUDOP - Wyszukiwanie pomocy otrzymanej przez beneficjenta, Mapa dotacji UE, Centralny Rejestr Pełnomocnictw Ogólnych, Krajowy Rejestr Zadłużonych, Decyzje Prezesa UOKiK, Wyroki dotyczące Decyzji Prezesa UOKiK, Europejski portal e-sprawiedliwość, Krajowy Rejestr Długów, Krajowe Biuro Informacji Gospodarczej, CRIF i inne
- nieustrukturyzowane dane ogólnodostępne, głównie w domenie publicznej, czyli z szeroko pojętego otoczenia internetowego. Wykorzystywane będą (głównie za pośrednictwem API lub w przypadku braku API, metodą wyodrębniania) m.in. następujące źródła danych: bazy WHOIS, strony www, bazy adresów IP, adresy e-mail (weryfikacja, wyszukiwanie, wycieki).

W ramach zadania Zleceniodawca oczekuje również dostarczenia wsparcia merytorycznego w procesie generowania lub gromadzenia danych badawczych w opracowanym schemacie danych włącznym wymiarze do 150 godzin roboczych. W szczególności w przypadku braku możliwości zgromadzenia danych rzeczywistych wyżej opisane wsparcie zostanie wykorzystane do wygenerowania danych syntetycznych (symulacyjnych).

Termin realizacji zadania: od 01.01.2024 do najpóźniej 31.12.2024 (maksymalnie 12 miesięcy), może ulec zmianie w zależności od daty rozpoczęcia projektu budowy Integratora.

Załącznik numer 3 do postępowania

W zależności od postępów prac oraz od ustalonego Harmonogramu świadczenia usługi realizacja Zadania A może być częściowo zrównoleglona z realizacją Zadania B, o którym mowa poniżej.

Etap 2 – Zadanie B – opracowanie algorytmów identyfikacji ryzyk i modelu scoringu

W wyniku realizacji zadania opracowana zostanie metoda połączenia praktycznie nieograniczonej liczby źródeł wiedzy o dowolnym charakterze i strukturze w celu usprawnienia procesu decyzyjnego danej organizacji. Dokonana zostanie identyfikacja sieci powiązań podmiotów i wynikających z tego ryzyk wraz z określeniem scoringu dla danego podmiotu. Kluczowym elementem będzie połączenie ogólnodostępnych rejestrów, baz danych, wykazów, danych statystycznych, raportów, które agregują informacje jawnoźródłowe. Integrator danych przy wykorzystaniu m.in. „fuzzy matchingu” wskaże powiązania, relacje między obiektami, nawet w przypadku braku oczywistej zależności między nimi, wynikającej np. z różnego rodzaju zapisu danych. Integrator umożliwi w ten sposób dostarczenie bieżących raportów o zagrożeniach, ich prawdopodobieństwie, potencjalnych skutkach i rozmiarach ewentualnych strat. W oparciu o przedstawione wyniki w raporcie końcowym, użytkownik otrzyma informacje wskazujące na potencjalne ryzyka wynikające z prowadzonej działalności gospodarczej, będącej przedmiotem analizy.

W ramach zadania powinna powstać również metoda oceny skuteczności opracowanych algorytmów scoringowych wraz z określeniem poziomu ryzyka w zakresie prowadzonej działalności gospodarczej.

W zależności od postępów prac oraz od ustalonego Harmonogramu świadczenia usługi realizacja Zadania B może być częściowo zrównoleglona z realizacją Zadania C, o którym mowa poniżej.

Termin realizacji zadania: najpóźniej do 31.12.2025, może ulec zmianie w zależności od daty rozpoczęcia projektu budowy Integratora oraz terminu realizacji Zadania A.

Etap 3 – Zadanie C – przeprowadzenie testów oraz optymalizacja algorytmów scoringowych

Zadanie polega na przeprowadzeniu testów opracowanych w Zadaniu B algorytmów i mechanizmów scoringowych na dostarczonej próbce danych rzeczywistych w odniesieniu do wybranych przez Zleceniodawcę podmiotów (od 1 do 5 podmiotów, w zależności od zapotrzebowania), a także na optymalizacji i na dostrojeniu i ewentualnej korekcji algorytmów w celu uzyskania optymalnych rezultatów w postaci trafności predykcji oraz efektywnego wykorzystania zasobów infrastrukturalnych (storage, compute). Do oceny trafności scoringu zostanie wykorzystana opracowana w Zadaniu B metoda oceny efektywności algorytmów.

Termin realizacji zadania: najpóźniej do 30.09.2026, może ulec zmianie w zależności od daty rozpoczęcia projektu budowy Integratora oraz terminu zakończenia realizacji Zadania B.

Harmonogram świadczenia usługi

Szczegółowy harmonogram prac w każdym z Etapów realizacji usługi ustalany będzie z Wykonawcą z miesięcznym wyprzedzeniem, na podstawie zapotrzebowania Zleceniodawcy i możliwości Wykonawcy, z uwzględnieniem terminów opisanych wyżej.



Załącznik numer 3 do postępowania

Jednocześnie ze względu na konieczność zapewnienia ciągłości i sprawnego przebiegu prac B+R w projekcie:

- usługa powinna być świadczona w sposób ciągły przez cały okres trwania wyżej opisanych Zadań,
- odbiór usługi i zapłata wynagrodzenia Wykonawcy będą się odbywać po zrealizowaniu każdego etapu i w oparciu o wytworzone w ramach zakończenia świadczenia usługi dokumenty techniczne i projektowe.

Miejsce realizacji zamówienia

Usługa będzie realizowana w siedzibie Zamawiającego bądź w innym miejscu uzgodnionym z Wykonawcą. Zamawiający dopuszcza również możliwość zdalnego świadczenia usługi.

Zamawiający na potrzeby realizacji zadań udostępni Wykonawcy niezbędny sprzęt komputerowy z oprogramowaniem i inną niezbędną infrastrukturę do wyłącznego zastosowania w ramach realizacji przedmiotu zamówienia.