

Załącznik nr 1 – opis przedmiotu zamówienia nr: Z01/10/2024/BCHANGE

Postępowanie ofertowe pn.	„Nabór na Podwykonawcę do realizacji badań przemysłowych w ramach realizacji projektu składanego przez BCHANGE Sp. z o.o., w konkursie – Ścieżka SMART”
----------------------------------	--

OPIS PRZEDMIOTU ZAMÓWIENIA

OPIS PROJEKTU

Przedmiotem Zamówienia będzie Wsparcie Zamawiającego w badaniach mających na celu stworzenie automatycznego systemu anotacji i augmentacji danych na potrzeby tworzonego systemu sztucznej inteligencji dla wsparcia klientów Zamawiającego w obsłudze zapytań z obszarów prawa pracy i obsługi pracowniczej.

Zakres prac badawczych i rozwojowych

Etap 1:

1. Przygotowanie danych. Podwykonawca będzie odpowiedzialny za przygotowanie danych treningowych i testowych. W ramach tego zadania Podwykonawca także wyszuka, wybierze i pozyska, w konsultacji i po akceptacji Zamawiającego, najbardziej adekwatne do zadania Źródła danych. Podwykonawca stworzy narzędzia do generowania danych, anotowania, wykrywania POS i NER oraz ich modyfikacji w sposób iteracyjny celem osiągnięcia zakładanych kamieni milowych Projektu. Zadaniem podwykonawcy będzie przygotowanie zbiorów danych. Podwykonawca będzie zobowiązany do zebrania i usystematyzowania danych z różnych Źródeł posiadanych przez Wnioskodawcę, takich jak bazy wiedzy, bazy danych, korpusy tekstowe oraz dane posiadane przez biznes. Zostaną wykonane wszelkie niezbędne czynności, takie jak weryfikacja jakości danych, czyszczenie i formatowanie.
2. Przetwarzanie danych - zadaniem Podwykonawcy będzie przetwarzanie danych w celu uzyskania odpowiednich struktur danych w odpowiednich formatach, które będą używane do trenowania modeli głębokich sieci neuronowych z atencją. Podwykonawca będzie musiał stworzyć, a następnie wykorzystać odpowiednie narzędzia umożliwiające uzyskanie danych treningowych w odpowiedniej formie wymaganej we wsadowym uczeniu modeli. Zakres niezbędnych przekształceń zawiera się, ale nie ogranicza się do następującej listy: usuwanie duplikatów, usuwanie niepotrzebnych danych, formatowanie, indeksowanie, rozmnażanie i zwielokrotnianie. W ramach tego zadania podwykonawca zobowiązany będzie także do stworzenia automatycznych przepływów (ang. pipeline) służących do czyszczenia danych poprzez m.in.: usuwanie zbędnych znaków, formatowanie tekstu, tak aby dane były spójne i jednorodne, przeprowadzenie tokenizacji, normalizacji tekstu. usunięcia tzw. stop-words oraz dokonanie lematyzacji.
3. Annotowanie danych - zadaniem podwykonawcy będzie oznaczenie elementów tekstu, takich jak np: nazwy własne, kategorie, tagi czy relacje między elementami. Podwykonawca zobowiązany będzie do dostarczenia narzędzia umożliwiającego masowe i efektywne annotowanie danych, wykorzystujące najnowsze osiągnięcia naukowe w zakresie semi-automatycznego annotowania danych i wsparcia

annotatorów poprzez wykorzystanie funkcji heurystycznych i programowania danych.

Etap 2:

1. Implementacja architektury. Zadaniem podwykonawcy będzie implementacja architektury opracowanej przez Zamawiającego oraz przeprowadzenie odpowiednich testów celem dobrania odpowiednich dla wybranej architektury hiperparametrów oraz wykorzystanie technik parametryzacji i RLHF.
2. Trenowanie modeli. Podwykonawca przeprowadzi trening na danych treningowych. Podczas trenowania modelu będzie odpowiedzialny za dostosowanie parametrów modelu oraz preprocessing danych pod wymagany format sieci. Podwykonawca odpowiedzialny będzie także za dostarczenie niezbędnej infrastruktury technicznej do przeprowadzenia procesu trenowania modelu. W tym celu należy odpowiednio dostosować hiperparametry i wykorzystać algorytm optymalizacji, tak aby model osiągnął jak najlepsze wyniki.
3. Testowanie, modyfikacja i walidacja modeli. Zadanie Podwykonawcy polegać będzie na przeprowadzeniu procesu testowania i optymalizacji modeli mając na uwadze zestaw danych testowych. W celu oceny skuteczności modelu wykorzystane będą metryki, takie jak precyzja, czułość czy specyficzność. Celem Podwykonawcy będzie osiągnięcie wartości parametrów określonych przez Zamawiającego spełniających określone wartości kamieni milowych. Wnioskodawca będzie zobowiązany do iteracyjnego dostosowywania parametrów modelu, powtarzania procesu trenowania i testowania modeli celem uzyskania zakładanych wartości brzegowych.
4. Podwykonawca będzie odpowiedzialny za wdrożenie modelu jako samodzielny komponent systemowy pozwalający na dalsze testy - w tym za zintegrowanie modelu z innymi narzędziami i systemami, takimi jak bazy danych czy platformy analizy danych oraz komponentami powstałymi w poprzednim etapie prac.

Etap 3

1. Operacjonalizacja modeli. W ramach tego Zadania, Podwykonawca wraz z Zamawiającym przeprowadzi proces operacjonalizacji opracowywanych modeli i dostosowania ich do warunków zbliżonych do produkcyjnych. Podwykonawca odpowiedzialny będzie za prace programistyczne, testy i prace związane z administracją systemami- w tym stworzenie skryptów Dockerfile, Docker-Compose i Terraform oraz przeprowadzenie testów bezpieczeństwa oraz testów wydajnościowych.
2. Stworzenie przepływów danych i akcji do dotrenowywania modeli - W ramach tego zadania, podwykonawca powinien stworzyć przepływy danych i akcji, które umożliwiają dotrenowanie modeli. Proces dotrenowania polegać będzie na dostosowaniu modelu do bieżących warunków, w których będzie operacjonalizowany. Przepływy danych i akcji powinny umożliwiać dostęp do danych treningowych i testowych, jak również do narzędzi i raportów, które mogą pomóc w procesie dotrenowania.
3. Stworzenie środowiska uruchomieniowego - W ramach tego zadania, podwykonawca powinien stworzyć środowisko uruchomieniowe, w którym będą działać opracowane modele. Środowisko to powinno zapewniać stabilną i wydajną platformę, na której będą działać modele. Powinno również umożliwiać łatwe zarządzanie i konfigurację modeli oraz monitorowanie ich działania.
4. Optymalizacja wydajnościowa systemu - W ramach tego zadania, podwykonawca powinien przeprowadzić optymalizację wydajnościową systemu. Optymalizacja ta powinna obejmować takie

czynności jak optymalizacja kodu, konfiguracja infrastruktury sprzętowej oraz optymalizacja procesów działania systemu. Celem tej optymalizacji jest uzyskanie jak najwyższej wydajności systemu przy jak najniższym obciążeniu sprzętowym.

5. Przetestowanie merytoryczne modułów stworzonych w ramach poprzednich Etapów. Testowanie to polegać będzie na przetestowaniu działania modułów i ocenie ich skuteczności w realizacji założonych celów. Testy te powinny obejmować takie czynności jak testowanie funkcjonalności, testowanie wydajności oraz testowanie odporności systemu na różnego rodzaju błędy i awarie.
6. Wykonanie testów integracyjnych, akceptacyjnych i wydajnościowych - W ramach tego zadania, podwykonawca powinien przeprowadzić testy integracyjne, akceptacyjne i wydajnościowe. W ramach tego zadania wykonane zostaną scenariusze testów akceptacyjnych w języku Gerkhin oraz stworzone zostaną odpowiednie skrypty umożliwiające przeprowadzenie zautomatyzowanych testów end-to-end. Testy te mają na celu sprawdzenie, czy opracowane moduły działają poprawnie wraz z innymi modułami systemu, czy spełniają wymagania użytkowników oraz czy są wystarczająco wydajne. Testy te powinny obejmować takie czynności jak testowanie integracji modułów, testowanie funkcjonalności w warunkach produkcyjnych oraz testowanie wydajności systemu pod różnymi warunkami obciążenia.