

Załącznik nr 4a

Opis koncepcji realizacji projektu pt. **Chmurowa platforma (CAISE) do wytwarzania uniwersalnych usług inteligentnych dla różnych obszarów zastosowań**

CAISE Platform = Cloud Artificial Intelligence Software Engineering Platform

1. Wykonawcą projektu jest Politechnika Gdańska (indirect partner), a w szczególności dwie jej jednostki:

- Centrum Informatyczne Trójmiejskiej Sieci Komputerowej – CI TASK
- Wydział Elektroniki Telekomunikacji i Informatyki – WETI

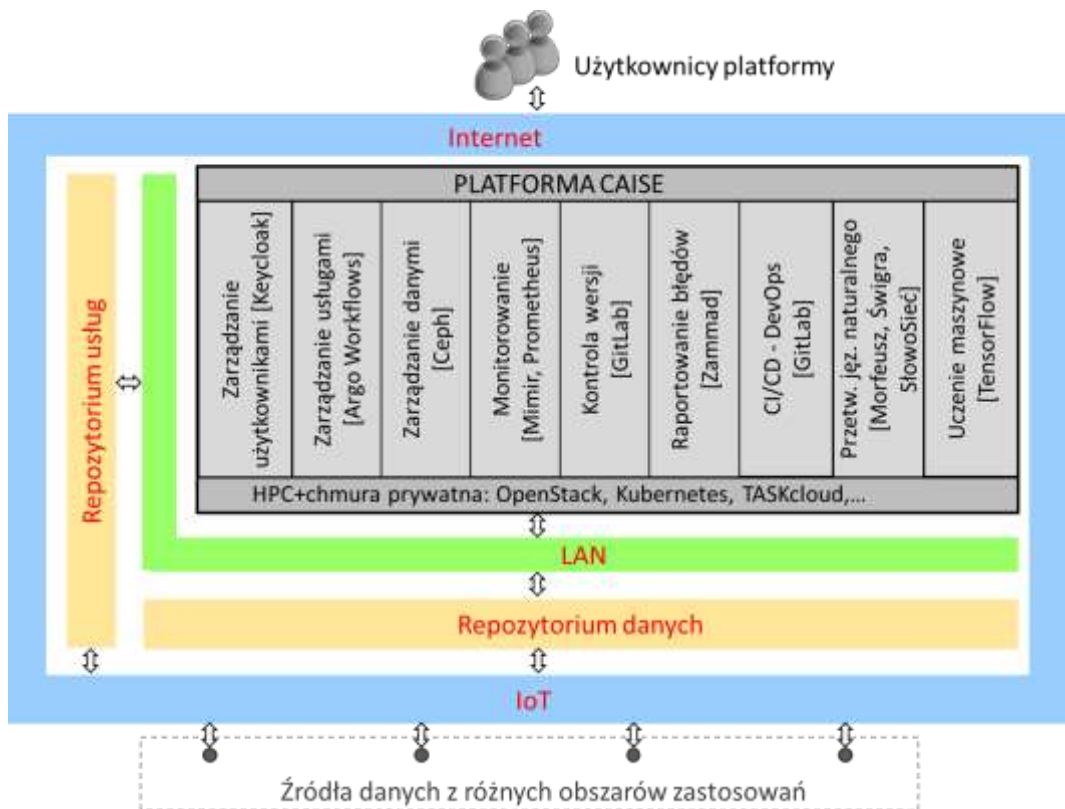
Dodatkowo:

- Podwykonawcy Projektu (3 krajowe firmy MŚP) wyłonieni z przetargu do testowania i rozwoju platformy CAISE poprzez ocenę jej możliwości w wybranych (różnych) sektorach działalności publicznej. Zakłada się udział Podwykonawców w projekcie CAISE po pierwszym roku jego realizacji, przez następne 1.5 roku, aż do końca projektu.
- Współpracujący z projektem (4 firmy zagraniczne – 2 węgierskie, 1 włoska i 1 hiszpańska) – podpisana deklaracja o współpracy w ramach programu UE – IPCEI-CIS.

2. Cel projektu CAISE

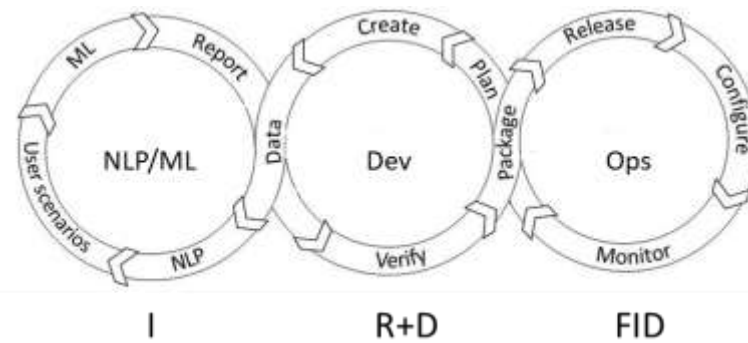
Celem projektu jest budowa innowacyjnej platformy chmurowej CAISE wspomagającej wytwarzanie usług analizy treści dokumentów cyfrowych w języku polskim przy wykorzystaniu NLP (Natural Language Processing) oraz uczenia maszynowego ML – Machine Learning) wymaganych przy budowie różnego typu usług i aplikacji sektorowych, na ogół wykorzystujące sekwencje działań takich usług. Koncepcja projektu jest przedstawiona na rys. 1. Platforma zostanie zaimplementowana na chmurze TASKcloud (zbudowana i eksploatowana przez CI TASK), przy wykorzystaniu różnego typu oprogramowania otwartego. Takie oprogramowanie jest rozwijane przez społeczności światowe i może być powszechnie wykorzystane.

Docelowo platforma CAISE będzie oferowała usługi wspomagające budowę różnych aplikacji, przydatnych do rozwoju cyfryzacji sektora działalności publicznej. Będą one, po zakończeniu projektu, implementowane przez zainteresowane firmy komercyjne, przy wykorzystaniu możliwości oferowanych przez platformę CAISE. Zaletami takiego podejścia są niższe koszty wytwarzania aplikacji poprzez skrócenie czasu wytwarzania oraz niepowielanie powtarzających się czynności przy budowie różnych aplikacji od podstaw, a także prostota ich tworzenia (inteligentne usługi wspomagające) mimo złożoności rozpatrywanych problemów. W procesie ciągłego testowania i udoskonalania platformy (techniki CI/CD (Continuous Integration/Continuous Delivery, patrz rys.2) istotną rolę odegrają podwykonawcy/firmy, które dzieląc się komercyjnym doświadczeniem, dostarczą odpowiednich danych rzeczywistych do rozwoju platformy, przetestują jej możliwości oraz zweryfikują spełnianie wymagań rynkowych.



Rys.1 Koncepcja projektu CAISE

W ramach rozwoju platformy utworzone zostanie repozytorium jej usług, a także repozytorium danych związanych z uczeniem modeli sieci neuronowych stanowiących główne komponenty takich usług. Dodatkowo w repozytorium danych znajdować się będą wyniki oceny platformy przez podwykonawców i innych klientów, dotyczące zarówno użyteczności platformy jak i jakości wytwarzanych usług. Zapewni to przyjęta cykliczna metoda wytwarzania DevOps przy wykorzystaniu techniki CI/CD. W procesie wytwarzania platformy istotną rolę odegrają dwie powiązane ze sobą kategorie cykli rozwojowych: technologiczne oraz funkcjonalne. Cykle technologiczne dotyczą rozwoju platformy od poziomu gotowości TRL3 do poziomu gotowości TRL 9. Cykle funkcjonalne, podrzędne do technologicznych, związane są z rozwojem poszczególnych funkcji platformy oraz usług chmurowych wraz z zapewnieniem wymaganego poziomu jakości.



Rys. 2 Zakładana metodyka prac w projekcie CAISE

Zaangażowanie zewnętrznych firm jako podwykonawców w proces cyklicznego testowania i rozwoju platformy zapewni uzyskanie przez nią najwyższego stopnia gotowości technologicznej, jak również zapewni osiągnięcie wysokiej użyteczności poprzez realizację rzeczywistych przypadków użycia (use cases) adekwatnych dla rozpatrywanych sektorów działalności publicznej: wymiaru sprawiedliwości, służby zdrowia oraz oświaty.

Istotą realizowanej działalności badawczo-rozwojowej jest zastosowanie i rozwój metodyki DevOps wytwarzania inteligentnych usług chmurowych, integracja różnych narzędzi otwartego oprogramowania wspomagających tego typu procesy, a także opracowanie modeli weryfikacji i oceny jakości platformy oraz wytwarzanych usług. Istotą wdrożenia jest sama platforma oraz rozbudowa repozytorium usług i danych wspomagających budowę różnego typu aplikacji związanych z ww. sektorami działalności publicznej.

3. Pojęcia kluczowe:

Platforma CAISE – główny produkt projektu CAISE, infrastruktura chmurowa na poziomie usług PaaS, wspomagająca zarówno rozwój usług inteligentnych z zakresu analizy treści dokumentów cyfrowych jak też budowę aplikacji sektorowych z wykorzystaniem takich usług. Platforma powstanie na bazie chmury TASKcloud dostępnej w CI TASK Politechnika Gdańska. Innowacyjność rozwiązania wynika z integracji nowoczesnych technologii chmury obliczeniowej, Big Data oraz sztucznej inteligencji (AI) wraz z wykorzystaniem obliczeń dużej skali (HPC). Poza tym proponowane rozwiązanie będzie uniwersalne z uwagi na wykorzystanie w różnych sektorach działalności ludzkiej, w których istotną rolę odgrywają dokumenty cyfrowe. Automatyczna analiza takich dokumentów oraz wyłuskiwanie zawartej tam wiedzy pozwoli na racjonalne podejmowanie decyzji oraz realizację wymaganych scenariuszy działań dla wielu powtarzających się czynności wielokrotnie realizowanych dotąd ręcznie przez człowieka.

Proponowane usługi chmurowe platformy CAISE zostaną zakwalifikowane do trzech **kategorii usług**:

- XaaS, gdzie $X \in \{I, P, S\}$; $X \in \{(I)nfrastructure as a Service, (P)latform as a Service, (S)oftware as a Service\}$ - to podstawowe warstwy usług chmury obliczeniowej związane z infrastrukturą fizyczną chmury TASKcloud (IaaS), infrastrukturą wspomagającą wytwarzanie usług i aplikacji na tej chmurze (PaaS) oraz z zapewnieniem dostępu do wytworzonych aplikacji (SaaS) oraz ich wykonanie w środowisku chmurowym.

- YaaS, gdzie $Y \in \{NLP, ML\}$; są to usługi przetwarzania dokumentów (tekstów i obrazów) cyfrowych tworzonych dla języka polskiego (NLP) wspomaganych usługami uczenia maszynowego (ML), tworzone z wykorzystaniem metod sztucznej inteligencji oraz ogólnie dostępnych zbiorów danych. Tego typu usługi zorientowane na konkretny sektor zastosowań oznaczono przez $Y(\text{sektor})aaS$.
- ZaaS, gdzie $Z \in \{Ma(nagement), Mo(nitoring), Q(uality), R(eporting)\} = \{Ma, Mo, Q, R\}$; są to usługi wspomagające funkcjonowanie platformy umożliwiające jej użytkownikom przyjazne i efektywne korzystanie z usług typu X i Y, a także wspierające procesy oceny jakości usług oraz użyteczności platformy. Tego typu usługi dotyczą wspomaganie zarządzania użytkownikami i danymi (Ma), jak również monitorowania i raportowania przebiegu realizacji procesów wytwarzania (Mo oraz R), a także oceny jakości modeli, usług i tworzonych na platformie aplikacji (Q).

Przypadki użycia (use cases) definiują wybrane sytuacje, mające miejsce w rozważanych obszarach działalności publicznej, które wymagają odpowiedniego postępowania, a które mogą być finalnie przedstawione jako realizacja wybranych usług Y, Z. Będą one określane na podstawie rzeczywiście wykonywanych czynności w konkretnym sektorze działalności człowieka. W naszym przypadku, każda sytuacja jest opisana poprzez odpowiednie dane (dokumenty) z nią związane, zaś działania wskazują na czynności niezbędne z ich analizą w celu określenia odpowiedniej relacji przyczynowo skutkowej. Przypadki użycia są niezbędne do testowania funkcjonalności oraz jakości działania platformy CAISE.

Metodyka DevOps – aktualna i popularna metodyka (kultura) wytwarzania platformy i chmurowej polegająca na ciągłym wykorzystaniu technik (CI/CD) umożliwiającej jej testowanie oraz badaniu jej użyteczności w różnych zastosowaniach przy wykorzystaniu użytkowników z różnych sektorów działalności publicznej. Polega na cyklicznym rozwoju platformy na dwóch poziomach: wyższym - cyklach technologicznych zapewniającym przejście do kolejnego poziomu gotowości technologicznej oraz niższym cyklach funkcjonalnych polegającym na rozbudowie usług platformy (jej funkcjonalności i jakości) na danym poziomie technologicznym.

Model oceny jakości – określają atrybuty, charakterystyki oraz metryki jakości oraz procedury pozwalające na wyznaczenie tych metryk jakości na podstawie ciągłego monitorowania procesu (CI/CD) wytwarzania platformy. Metryki te będą śledzone i analizowane na bieżąco (usługi typu Z) w celu oceny postępów rozwoju platformy na podstawie oceny jej zachowania się dla różnych przypadków użycia. Podstawowe metryki jakości to: funkcjonalność – zakres funkcji platformy dostępnych użytkownikom, wydajność platformy określana przez średnie czasy wykonania poszczególnych usług na tej platformie, użyteczność platformy jako stosunek liczby usług wykorzystywanych do liczby wszystkich oferowanych usług Y, Z oferowanych na platformie, wiarygodność platformy jako stosunek liczby wyeliminowanych problemów do liczby zgłaszanych problemów podczas rozwoju platformy oraz dojrzałość platformy w zadanym czasie jako stosunek liczby dostępnych usług w poprzednim cyklu wytwarzania do liczby wszystkich usług dostępnych i nowo utworzonych w następnym cyklu, jak również przyjazność wykorzystywanych interfejsów platformy.

Zakres prac dotyczący jednostek PG (CI TASK, WETI) oraz firm/podwykonawców jest wskazany w kolejnych punktach.

4. Zadaniem CI TASK PG jest utrzymanie, aktualizacja i rozwój środowiska chmurowego TASKcloud do potrzeb projektu, a w szczególności:

- wdrożenie platform OpenStack oraz Kubernetes dla realizacji usług na poziomie IaaS (infrastruktury sprzętowej) oraz PaaS (usługi systemowe), zapewniających efektywne odwzorowanie maszyn wirtualnych oraz kontenerów na platformę fizyczną chmury obliczeniowej minimalizujących zużycie energii oraz skalowalność obliczeń, a także bezpieczeństwo przetwarzania.

- wdrożenie platform zarządzania kodem i wsparcia metodyki DevOps (np. GitLab) oraz systemu obsługi zgłoszeń, (np. Zammad) do kontroli wersji tworzonych usług oraz monitorowania błędów oprogramowania, a także wsparcia wykorzystania techniki CI/CD DevOps przy realizacji platformy.

- wdrożenie platformy monitoringu (np. Prometheus, Mimir) do monitorowania uruchomionych usług procesu wytwarzania usług oraz platformy do budowy scenariuszy usług metodyką DevOps, (np. Argo Workflows).

- wdrożenie platformy dla wspomagania wykorzystania metod sztucznej inteligencji (AI) oraz uczenia maszynowego (ML) do budowy modeli sieci neuronowych (np. TensorFlow).

- integracja współdziałania powyższych platform na poziomie funkcjonalnym w formie usług XaaS, $X=\{I,P,S\}$ oraz w formie usług ZaaS, $Z=\{Ma, Mo, R, Q\}$, umieszczonych w repozytorium usług, a także opracowanie poradnika użytkownika platformy CAISE.

- prowadzenie szkoleń z użytkowania tej platformy dla zainteresowanych użytkowników.

3. Zadaniem WETI PG jest przygotowanie algorytmów i oprogramowania (w formie bibliotek) analizy treści (tekstu i obrazów) dokumentów cyfrowych do realizacji usług na poziomie PaaS (usługi typu smart), w szczególności:

- opracowanie modeli dokumentów cyfrowych, metod reprezentacji tekstu i obrazów oraz procedur gromadzenia skalowalnych wolumenów dokumentów niezbędnych do budowy modeli uczących dla zadań segmentacji, ekstrakcji i klasyfikacji ich treści.

- rozwój metod i algorytmów wydobywania pojęć z tekstów nieustrukturalizowanych (ang. text mining) w języku polskim i automatycznej analizy ich znaczenia, w tym również procedur animizacji dokumentów.

- przygotowanie bibliotek uniwersalnych procedur analizy dokumentów w języku Python dla różnych obszarów zastosowań oraz opracowanie na platformie CAISE chmurowych usług bazowych typu YaaS wraz z dokumentacją ich wykorzystania,

- przygotowanie repozytorium usług YaaS oraz danych uczących do dalszego rozwoju platformy CAISE zorientowanych na dany sektor zastosowań,

- prowadzenie szkoleń z zakresu metod inteligentnej analizy treści dokumentów dla firm korzystających z platformy CAISE do budowy nowych usług typu YaaS.

4. Zadaniem wspólnym CI TASK i WETI – nieustanne śledzenie i korekta pojawiających się problemów przy budowie usług XaaS, YaaS, ZaaS, oraz platformy CAISE, jak również :

- przeprowadzenie procedury przetargowej na wyłonienie podwykonawców - firmy opracowujące i wdrażające usługi sektorowe typu YaaS w trzech wybranych obszarach sektora publicznego: wymiaru sprawiedliwości, ochrony zdrowia oraz oświaty,

- monitorowanie zadań wykonywanych przez firmy wg założonego harmonogramu i kontrola postępu ich prac w oparciu o uzgodnione kamienie milowe (ang. milestones) i wartości mierników poziomu gotowości technologicznej (TRL – Technology Readiness Level), a także ciągłe udoskonalanie platformy CAISE oraz wytwarzanych usług typu Y i Z, zapewniając osiągnięcie najwyższego poziomu gotowości technologicznej.

5. Zadania podwykonawców dotyczą rozwoju usług typu YaaS oraz oceny jakościowej platformy CAISE, zakłada się wykorzystanie 3 etatów z każdej firmy do realizacji zadań projektowych, w szczególności:

- wykorzystanie platformy CAISE do budowy usług typu $Y(\text{sektor})=Y(\text{sektor})aaS$, gdzie $\text{sektor} \in \{\text{Wymiar sprawiedliwości, Ochrona zdrowia, Oświata}\}$; bazując na usługach YaaS. W praktyce oznacza to wykorzystanie metody tworzenia usług YaaS z uwzględnieniem zaproponowanego przez firmy zbioru danych: uczących oraz sprawdzających. Taki zbiór danych dostarczy każdy z Podwykonawców z własnego sektora zastosowań.

- realizacja na platformie CAISE przypadków użycia wskazanych dla każdego z sektorów działalności publicznej oraz wykorzystujących dostępne usługi platformy CAISE oraz usługi $Y(\text{sektor})aaS$. Celem wykonania tych scenariuszy jest sprawdzenie zarówno poprawności działania realizowanych usług jak i zadanej funkcjonalności oraz jakości tej platformy, a także metodyki jej wytwarzania.

- wykorzystanie usług ZaaS do monitorowania przebiegu procesów wytwarzania usług i realizacji przypadków użycia oraz raportowanie danych niezbędnych do ocen jakości platformy CAISE oraz udział w ciągłym rozwoju platformy poprzez monitorowanie i sygnalizowanie występujących anomalii w celu uzyskania przez Wykonawcę akceptowalności poziomu gotowości technologicznej TRL 9.

- sporządzanie raportów z rozwoju platformy CAISE w każdym cyklu technologicznym oraz funkcjonalnym wraz ze wskazaniem zmian jej metryk jakościowych.

6. Opis etapów – WP (Work Packages)

Wytwarzanie usług i aplikacji będzie realizowane w trybie ciągłym zgodnie z metodyką DevOps+AI przedstawioną na rysunku 2. Jej kluczowe kroki w cyklu I (innovation) będą obejmowały analizę NLP nieustrukturyzowanych danych tekstowych (treści dokumentów) pochodzących z repozytoriów ogólnie dostępnych oraz użytkowników końcowych sektora publicznego, budowanie, prototypowanie i optymalizację modeli uczenia maszynowego (ML) do realizacji rzeczywistych (złożonych) przypadków użycia implementowanych na platformie CAISE przez Wykonawcę oraz firmy/podwykonawców. W cyklu R+D (Research and Development) następuje ciągły rozwój platformy poprzez rozbudowę usług Y,Z, natomiast w cyklu FID (First Industrial Deployment) proces wytwarzania jest raportowany, a wyniki

oceny dokonywanych zmian będą stanowiły podstawę do oceny dojrzałości technologicznej platformy. W projekcie zakłada się realizację 3 głównych zadań (WP – Work Packages) (patrz rys. 3):

WP1. Opracowanie koncepcji platformy i jej eksperymentalna weryfikacja

Prace etapu WP1 będą realizowane przez Wykonawcę i będą miały na celu ustawiczną integrację własnych zasobów chmury obliczeniowej TASKcloud dotyczących HPC (własność Wykonawcy projektu) i dostępnych otwarto źródłowych narzędzi i bibliotek do postaci platformy chmurowej (usługi XaaS) przedstawionej na Rysunku 1 oraz ciągłą aktualizację i utrzymanie oprogramowania w czasie trwania projektu oraz następnie w okresie jego trwałości.

WP2: Wytworzenie prototypu platformy i jej usług

Rozwój prototypu platformy od osiągniętego przez nią, na zakończenie etapu WP1 poziomu dojrzałości TRL-6, do docelowego poziomu TRL-8 umożliwiającego jej pierwsze wdrożenie przemysłowe. Będzie realizowane przez Wykonawcę projektu przy współpracy z trzema firmami podwykonawców. Umożliwi to zarówno udoskonalanie usług Y, Z, a także wytwarzanie usług Y(sektor) oraz testowanie możliwości platformy w założonym zakresie.

WP3: Walidacja¹ i ciągły rozwój platformy w środowisku docelowym;

Usługi YaaS i Y(sektor) oraz ZaaS platformy CAISE, rozwijane od etapu WP2 w trybie ciągłym CI/DC (model DevOps rozwoju oprogramowania), wykorzystane zostaną do budowy przypadków użycia, sprawdzających w warunkach operacyjnych (rynkowych) jej funkcjonalność i użyteczność. Rozwój ten zapewni jej wdrożenie w docelowym (rzeczywistym) środowisku operacyjnym firm podwykonawców (po osiągnięciu kamienia milowego M24 na zakończenie etapu WP2). Ponieważ jej usługi zostaną wykorzystane w procesie wytwarzania przez trzy krajowe firmy MŚP w trzech różnych segmentach sektora publicznego (wymiarze sprawiedliwości, ochronie zdrowia i oświacie) to przeprowadzona walidacja umożliwi przeprowadzenie pierwszego wdrożenia przemysłowego platformy CAISE (kamień milowy M30, patrz rys. 3).

7. Harmonogram działań projektu CAISE przedstawia poniższy wykres Gantta:

¹ Walidacja to potwierdzenie, że spełnione zostały wymagania odnośnie konkretnego zastosowania. Jest szeroko stosowana w technice i systemach zarządzania jakością. Może dotyczyć metody badawczej, procesu, procesu pomiarowego i środków kontroli. Proces walidacji obejmuje ustalenie wymagań, zestawienie procesu, określenie właściwości, porównanie z wymaganiami i generowanie certyfikatu (cyt. z Encyklopedia zarządzania, <https://mfiles.pl/pl/index.php/Walidacja>)

