

Program Strategiczny INFOSTRATEG „Zaawansowane technologie informacyjne, telekomunikacyjne i mechatroniczne”

Warszawa, 22 kwietnia 2020 r.

Spis treści

1. Wprowadzenie.....	3
2. Diagnoza sytuacji w obszarach nauki i gospodarki, które mają być objęte Programem .	4
Diagnostyka medyczna.....	5
Rozpoznawanie i analiza obrazów	7
Automatyzacja i robotyzacja rolnictwa.....	8
Automatyzacja obsługi klienta/pacjenta.....	9
Analiza treści publikowanych w Internecie	11
Blockchain	11
Analiza warunków prowadzenia prac B&R w ICT.....	11
Analiza potencjału B&R w dziedzinie sztucznej inteligencji	12
Analiza zgodności z dokumentami strategicznymi.....	13
3. Cel główny i cele szczegółowe	14
4. Zakres tematyczny Programu	16
T1) Inteligentny system przetwarzania mowy dla lekarzy.....	17
T2) Rozpoznawanie obrazów medycznych	18
T3) Diagnostyka wrodzonych wad metabolizmu w badaniach przesiewowych noworodków.....	19
T4) Różne scenariusze rozpoznawania zdjęć satelitarnych i lotniczych	20
T5) Inteligentna maszyna do zbierania jabłek	21
T6) Scenariusze selektywnej ochrony roślin.....	23
T7) Inteligentny dyspozytor korespondencji	24
T8) Asystent petenta w instytucjach użyteczności publicznej	25
T9) Wydajny i skalowalny blockchain konsorcyjny dla inteligentnych kontraktów	25
T10) Rozpoznawanie na filmie zachowań ludzkich	27
T11) Weryfikowanie źródeł informacji i detekcja fake newsów	28
T12) Analiza polskich danych medycznych.....	29
5. Sposób interwencji i warunki realizacji projektów w ramach Programu	30
5.1. Konkursy.....	32
6. Sposób monitorowania i oceny realizacji celów Programu	33

6.1. Ewaluacja	36
7. Określenie ryzyka dla osiągnięcia celów Programu	36
8. Harmonogram realizacji Programu	38
9. Budżet i plan finansowy Programu oraz źródła finansowania	38
10. System realizacji i zarządzanie Programem	39
11. Matryca logiczna Programu	41

1. Wprowadzenie

W ostatnich latach, w wyniku rozwinięcia nowoczesnych narzędzi i metod uczenia maszynowego (ang. *machine learning*), nastąpiła rewolucja w sztucznej inteligencji. Do popularyzacji SI przyczynił się też znacząco: dynamiczny rozwój łatwo dostępnych mocy obliczeniowych, wykorzystujących karty graficzne (ang. Graphic Processing Unit – GPU np. producenta NVIDIA), dostępność mocy obliczeniowych w chmurze (ang. *cloud computing*), zwiększona dostępność cyfrowych danych oraz popularyzacja otwartych narzędzi programistycznych (np. PyTorch, TensorFlow). Zmiany te przyczyniły się do rozwoju głębokiego uczenia sieci neuronowych (ang. *deep learning*), a także do dalszego rozwoju statystycznych metod uczenia maszynowego. Metody te i dostępne nowoczesne narzędzia pozwoliły na stworzenie algorytmów, które rozwiązują wiele problemów lepiej niż człowiek. Dla przykładu AlphaGo firmy Deepmind pokonało człowieka w grę Go, mimo iż wydawało się, że gra ta jeszcze długo pozostanie domeną człowieka. Dlatego naturalne stają się przewidywania, że uczenie maszynowe doprowadzi do stworzenia rozwiązań, które będą coraz bardziej obecne w naszym życiu.

Celem Programu INFOSTRATEG jest rozwijanie polskiego potencjału badawczego nad wybranymi problemami uczenia maszynowego, które mają potencjał rozwoju na szerszą skalę. Wybrane tematy badawcze są bardzo konkretne. Z jednej strony ma to umożliwić praktyczne zastosowanie stworzonych rozwiązań. Z drugiej strony koncentracja naszych starań na kilku wybranych tematach ma umożliwić wewnętrzną konkurencję, która doprowadzi do stworzenia doskonalszych rozwiązań. Jednocześnie warto podkreślić, że mimo, iż tematy te zostały sformułowane bardzo konkretnie, ich rozwiązanie wymaga stworzenia podstawowych technologii i rozwiązań, które znajdą w przyszłości szersze zastosowanie w gospodarce i nauce. Na pewno technologie wytworzone w pracy nad robotem zbierającym jabłka przydadzą się w pracy nad robotami dla innych zastosowań, a ograniczenie się do tego właśnie zastosowania pozwoli najpierw na stworzenie dobrego rozwiązania dla tego typu zastosowania w rolnictwie. Do działań NCBR w tym obszarze odniesiono się w *Polityce Rozwoju Sztucznej Inteligencji w Polsce na lata 2019 – 2027*, skierowanej do konsultacji społecznych w sierpniu 2019 r. W rozdziale dotyczącym ekosystemu AI i strategicznych czynników budowania potencjału AI w Polsce, program INFOSTRATEG wskazano jako potencjalne źródło finansowania i opracowania praktycznych zastosowań sztucznej inteligencji. *Polityka Rozwoju Sztucznej Inteligencji* wskazuje także na dwa kluczowe aspekty konieczne do skutecznego rozwoju SI w Polsce, do których w efektywny sposób przyczyni się program INFOSTRATEG. Jest to otwarcie i udostępnienie zbiorów danych. A także rozwój kompetencji w SI, np. poprzez wspieranie realizacji doktoratów wdrożeniowych w dziedzinie sztucznej inteligencji.

Oprócz uczenia maszynowego, technologia Blockchain w ostatnich latach, umożliwiła realizację pomysłów, które wcześniej występowały jedynie w literaturze *science fiction* oraz doprowadziła do powstania *eWalut*. Technologia ta ma szansę zostać szerzej zastosowana. Potencjalnie umożliwi ona zawieranie *eKontraktów*, ułatwiających współpracę między przedsiębiorstwami, poprzez

sformalizowanie warunków kontraktu i automatyzację ich egzekwowania. Dlatego właśnie ten temat, pomimo, że nie przewiduje wykorzystania uczenia maszynowego, został włączony w zakres tematyczny Programu INFOSTRATEG.

2. Diagnoza sytuacji w obszarach nauki i gospodarki, które mają być objęte Programem

Zastosowane podejście do diagnozy.

Poniższa diagnoza opiera się na analizach opracowanych na potrzeby Programu INFOSTRATEG przez zewnętrznych ekspertów, a także zawiera analizy przeprowadzone w ramach prac zespołu redakcyjnego.

Obszar objęty Programem.

Dbając o wysoką jakość i użyteczność rozwiązań jakie mają powstać w ramach INFOSTRATEGa, program będzie prowadzony m.in. w formule promującej wewnętrzną konkurencję i skoncentruje się na kilku wybranych tematach. Biorąc pod uwagę:

- 1) Zgłoszone przez interesariuszy zainteresowanie poszczególnymi tematami programu.
- 2) Fakt, że istnieją branże, których nie stać na prowadzenie prace B&R w obszarze SI, co grozi ich znacznym opóźnieniem w tym zakresie w stosunku do branż, które już realizują prace badawcze dot. SI
- 3) Brak uzyskania od przedstawicieli branż biorących udział w konsultacjach informacji zwrotnej co do konkretnych propozycji tematów programu.

Programem objęte będą zakresy tematyczne takie jak: medycyna, analiza obrazów, rolnictwo, administracja państwowa oraz technologia blockchain.

Wyniki analiz.

Przeprowadzone analizy wskazują, że jednym z głównych problemów, stojących na drodze szerszego rozwoju i wdrażania metod uczenia maszynowego, jest niedostateczny poziom organizacji (np. w zakresie kategoryzacji, opisanie czy oczyszczenia), integracji i analizy danych, uniemożliwiających ich pełne i optymalne wykorzystanie w procesie wytrenowania sztucznej inteligencji. Ten aspekt jest wyraźnie widoczny w przypadku zastosowań w medycynie, gdzie potencjalnie wiele informacji może być wykorzystanych do diagnostyki stanu pacjenta, prewencji oraz planowania leczenia. Tak się jednak nie dzieje w Polsce na dużą skalę z uwagi m.in. na brak dobrze opisanych danych medycznych, dostępnych w formie cyfrowej. Rozwijający się sektor technologii informacyjno-telekomunikacyjnych (ang. *ICT*) ze względu na wysoki stopień postępu technicznego nabiera coraz większego znaczenia w opiece zdrowotnej. Istnieje rosnące

zapotrzebowanie na technologie ICT dostosowane do uwarunkowań nowoczesnego systemu opieki zdrowotnej. Jednakże zastosowania medyczne ICT to tylko jeden z nielicznych przykładów. Wydaje się jasne, że nowoczesne metody uczenia maszynowego zaczną niedługo być obecne w coraz szerszym spektrum zastosowań. Obecnie sztuczna inteligencja w biznesie jest szeroko stosowana w automatyzacji marketingu i sprzedaży (np. profilowanie), obsłudze klienta (np. chat-boty), rozpoznawaniu twarzy i rzeczy (w tym w autonomicznych pojazdach) czy również w przetwarzaniu mowy i tekstu na potrzeby np. zautomatyzowanego tłumaczenia.

To wielka szansa, ale i wyzwanie dla polskich innowacyjnych firm stawiających na badania i rozwój w obszarze technologii ICT. Zmianie ulegają nie tylko metody analizy danych, ale także sztuczna inteligencja, która będzie w stanie podejmować/sugerować decyzje, a także samodzielnie sterować procesami, czy autonomicznymi pojazdami. Nie powstanie ona jednak bez dobrze opisanych cyfrowych danych, które posłużą do wytrenowania algorytmów. Jednocześnie eksperci wskazują na słabość B+R w Polsce w tej dziedzinie i koncentrację na badaniach "klasycznych" metod sztucznej inteligencji, które można też nazwać bazującymi na modelach (ang. *model-based*). Problem ten jest wyraźnie widoczny w robotyce, która przeżywa aktualnie ogromny rozwój dzięki wykorzystaniu głębokich sieci neuronowych. Aby ta wielka szansa mogła się zrealizować konieczne jest stymulowanie rozwoju B+R, wykorzystującego nowoczesne metody uczenia maszynowego w Polsce.

Poniżej wskazane obszary to dziedziny, w których uczenie maszynowe jest już wykorzystywane i zdaniem ekspertów zyska na znaczeniu w najbliższych latach.

Diagnostyka medyczna

Wśród najważniejszych narzędzi diagnostycznych we współczesnej medycynie są, uzyskiwane w różny sposób, obrazy narządów i tkanek. Obrazy powstają między innymi w wyniku prześwietlania promieniami Roentgena, obrazowania tomografem komputerowym, przy pomocy rezonansu magnetycznego lub mikroskopowej fotografii spreparowanej tkanki pobranej od pacjenta. Analiza tych obrazów jest czasochłonna oraz musi być przeprowadzana przez wysoko kwalifikowany specjalistyczny personel, którego w Polsce zaczyna coraz bardziej brakować.

Z drugiej strony oparte na uczeniu maszynowym algorytmy analizy obrazu są coraz skuteczniejsze i mogą usprawniać pracę specjalistów (patologów), w niektórych sytuacjach nawet ich wyręczając. Wiodące firmy, produkujące aparaturę medyczną, dołączają do niej elementy sztucznej inteligencji. Powstają też startupy testujące możliwość budowania dedykowanych narzędzi do analizy obrazów ustalonego typu.

W ciągu ostatnich dwudziestu lat tempo rozwoju najnowocześniejszych narzędzi do obrazowania medycznego było ogromne. Analiza i interpretacja danych, które się przy ich pomocy uzyskuje, wymaga ekspertów o coraz wyższych kompetencjach. Coraz trudniej jest też uzyskać informację

w pełni wykorzystać. W związku z tym w najbliższych kilkunastu latach na całym świecie będą rozwijane systemy do analizy obrazów medycznych. Jednakże możliwości oferowanych i testowanych dotychczas rozwiązań dalekie są od uniwersalności i elastyczności. Stworzenie generycznych rozwiązań uczenia maszynowego, umożliwiających analizę obrazów medycznych, umożliwiłoby dalszy rozwój i coraz szersze zastosowanie technik obrazowania w medycynie.

Kolejnym ważnym aspektem współczesnej medycyny jest prowadzenie badań na całych populacjach. W takim przypadku przeprowadzenie nawet stosunkowo prostych analiz stanowi duże wyzwanie, ze względu na liczebność zbioru pacjentów. Prowadzone w Polsce populacyjne badania przesiewowe u noworodków w kierunku chorób rzadkich pozwalają na wykrycie chorób w fazie przed objawowej. Prowadzone badania przesiewowe noworodków obejmują takie choroby, jak: fenyloketonuria, wrodzona niedoczynność tarczycy, mukowiscydoza, głuchota oraz od niedawna również rzadkie wrodzone wady metabolizmu. Dzięki staraniom Instytutu Matki i Dziecka wszystkie noworodki w Polsce poddawane są badaniom przesiewowym wykrywającym choroby, które w przypadku braku leczenia lub nawet leczenia spóźnionego o kilka dni (około tygodnia) grożą śmiercią lub trwałym kalectwem. Badania polegają na poddaniu próbek krwi lub moczu pacjenta analizie przy pomocy chromatografii gazowej i spektrometrii mas. W wyniku pomiaru powstają profile, których interpretacja wymaga czasochłonnych ocen przez wykształconych specjalistów. Dostępne narzędzia sztucznej inteligencji mogłyby doprowadzić do zbudowania algorytmów zastępujących ekspertów, co umożliwiłoby skrócenie czasu oczekiwania na diagnozę oraz rozszerzenie zakresu badań.

Nawet jeżeli nie prowadzimy celowych badań na całej populacji, to i tak coraz więcej danych medycznych o całej populacji jest dostępnych. W Polsce w tym momencie zbierane są w postaci cyfrowej dane chociażby o:

- wystawionych receptach;
- zrealizowanych receptach i sprzedanych lekach;
- świadczeniach medycznych zrealizowanych w ramach NFZ;
- zwolnieniach lekarskich.

Dostępność tego typu danych dla całej populacji pozwala na przeprowadzenie zaawansowanych analiz statystycznych, które mogłyby pozwolić chociażby na ocenę skuteczności różnych procedur medycznych, czy leków. To umożliwiłoby także prognozowanie całkowitych kosztów leczenia, a co za tym idzie podejmowanie bardziej korzystnych decyzji. Dane te pozwolą też na lepsze przewidywanie terminów, kiedy dana osoba powinna wykonać określone badanie profilaktyczne. Istotnym zastosowaniem tych informacji może być także wykrywanie nadużyć. Tego typu analizy mogą zostać wykonane przy pomocy statystycznych metod uczenia maszynowego.

Rozpoznawanie i analiza obrazów

Uczenie maszynowe, a w szczególności sieci neuronowe, w ostatnich latach zostało z sukcesem zastosowane do rozpoznawania obrazów, np. rozpoznawania twarzy. Jest też bardzo użyteczne w rozpoznawaniu i analizie zdjęć satelitarnych. Zdjęcia satelitarne potencjalnie dają możliwość automatycznego uzyskiwania różnego rodzaju informacji geolokalizacyjnej. Obecnie w wielu obszarach życia buduje się, rozwija i ulepsza systemy automatycznej analizy zdjęć satelitarnych. Brakuje jednakże narzędzi, które w łatwy sposób pozwalałyby na rozpoznawanie nowych typów obiektów, czy też obiektów o różnym charakterze. Dostępne narzędzia przystosowywane są do rozpoznawania jednego typu obiektu. Bez dogłębnej wiedzy o technikach uczenia maszynowego brakuje narzędzi, pozwalających na przeprowadzanie nowych analiz (każdy nowy przypadek użycia wymaga wiedzy eksperckiej). Problem nie ogranicza się jedynie do braku generycznych rozwiązań, a również do braku danych testowych, na których podstawie takie rozwiązania byłyby możliwe do zbudowania.

Warto zaznaczyć, że na świecie od niedawna powstają start-upy, które mają na celu stworzenie tego typu narzędzi np. SpaceKnow. Wciąż jednak wydaje się, że możliwości rozwiązań dalekie są jeszcze od osiągnięcia pełnej uniwersalności. Stworzenie tego typu uniwersalnego systemu stanowi więc ważne wyzwanie badawcze.

Flagowym zastosowaniem sieci neuronowych jest ww. rozpoznawanie obrazów. Jednakże sukcesy tych metod ograniczają się póki co jedynie do statycznych obrazów. Ogromnym wyzwaniem badawczym w tej dziedzinie jest wykorzystywanie tego typu technik do filmów. Dla przykładu: dość dobrze jesteśmy w stanie rozpoznawać jaką osobą znajduje się na zdjęciu, jednakże rozpoznanie co dana osoba robi na filmie jest już bardzo trudne. Odróżnienie czy dana osoba idzie, czy stoi, czy spożywa jabłko, czy też tylko je ogląda, wymaga analizy sekwencji wielu obrazów i wnioskowania bardziej kontekstowego. Wydaje się, że w najbliższych latach w tej dziedzinie może dokonać się przełom i powstaną dobrze działające techniki, pozwalające na rozpoznanie co dana osoba, bądź też grupa osób robi na filmie.

Automatyzacja i robotyzacja rolnictwa

Rolnictwo jest jednym z kluczowych obszarów polskiej gospodarki. Wprawdzie wartość surowców rolniczych to blisko 3% PKB, ale bez tych surowców nie może funkcjonować przemysł rolno-spożywczy, który jest największym przemysłem w Polsce. Produkty pochodzenia rolniczego stanowią blisko 13% polskiego eksportu. Wieś w Polsce jest ważna również z powodów społecznych, gdyż na terenach wiejskich mieszka blisko 40% Polaków. Na tle krajów Europy zachodniej polskie rolnictwo cechuje mniejsza wydajność produkcyjna i uprzemysłowienie produkcji oraz większa dekoncentracja gospodarstw rolnych, co utrudnia osiągnięcie efektu skali w kosztach produkcji. Obecnie polskie rolnictwo traci przewagi komparatywne związane z niskimi kosztami pracy. Rolnicy coraz częściej zgłaszają problemy ze znalezieniem pracowników, co przełoży się zapewne na wyższe

koszty produkcji rolnej. W tym samym czasie rolnictwo na świecie staje się coraz bardziej uprzemysłowione i zautomatyzowane. W USA wdrażane są już np. zautomatyzowane i bezobsługowe traktory oraz kombajny. Przy braku podjęcia zdecydowanych działań w zakresie zautomatyzowania, polskie rolnictwo może trwale utracić przewagę konkurencyjną. Z uwagi na fakt, iż Polska jest znaczącym eksporterem owoców i warzyw, jednym z ważniejszych kierunków produkcji rolnej w Polsce, wymagającym automatyzacji, są zbiory oraz przetwórstwo owoców i warzyw. Potrzeby w zakresie automatyzacji i robotyzacji są także dostrzegane w innych pracochłonnych kierunkach, jak choćby produkcja zwierzęca.

Na świecie rozwija się obecnie trend automatyzacji rolnictwa z wykorzystaniem sztucznej inteligencji oraz rozwiązań ICT, mających wspierać produkcję rolną. Sztuczna inteligencja znajduje zastosowanie nie tylko w kierowaniu pojazdami (np. traktory), ale też w: niszczeniu chwastów, diagnozie etapu rozwoju roślin, wskazaniu choroby i rodzaju bakterii, monitorowaniu stanu gleby, wykrywaniu rui u zwierząt, kontroli żywienia i zdrowia zwierząt oraz zapewnieniu im dobrostanu. W ostatnich latach na świecie powstają pierwsze firmy, które oferują inteligentne opryskiwacze, niosące środki ochrony roślin na wybrane rośliny (spółka Blue River Technology). Podejmowane są też próby opracowania robotów do zbioru jabłek konsumpcyjnych (spółka AbundantRobotics). Wykorzystanie sztucznej inteligencji w rolnictwie nie jest jeszcze masowe, jednak można o nim powiedzieć, że jest w bardzo wczesnej fazie rozwoju. Również w Polsce pojawiają się pierwsze wdrożenia rozwiązań opartych na sztucznej inteligencji. Przykładowo jeden z polskich producentów kwiatów ciętych w roku 2018 zaczął wykorzystywać nowoczesne produkty ICT do określenia etapu rozwoju rośliny. Niestety są to nadal wyjątki na rynku krajowym.

Powszechną techniką przy stosowaniu nawozów, środków ochrony roślin, a także przy nawadnianiu jest równomierne stosowanie takich samych dawek (na ogół maksymalnych zalecanych) dla całych plantacji. Lepszym rozwiązaniem jest dostosowanie dawki agrochemikaliów (środki ochrony roślin, nawozy) do potrzeb określonego fragmentu pola zgodnie z zasadami Rolnictwa Precyzyjnego, co w oczywisty sposób prowadzi do oszczędności i ograniczenia zagrożeń dla środowiska. Takie rozwiązanie wymaga jednak informacji o tym, jakich dawek środków ochrony roślin i nawozów należy użyć na danym fragmencie pola oraz przy jakich parametrach roboczych. Takie informacje mogą być zbierane w trakcie zabiegów przy użyciu układów identyfikacji (np. wielkość i gęstość roślin, obecność chwastów), a zróżnicowane dawki agrochemikaliów mogą być stosowane w czasie rzeczywistym. Ponadto parametry robocze inteligentnych maszyn mogą zmieniać się stosownie do lokalnych warunków (np. wielkość kropel, głębokość pracy agregatów ścierniskowych). W wielu miejscach na świecie prowadzone są badania na ten temat oraz powstają prototypy odpowiednich narzędzi, co świadczy z pewnością o dużym znaczeniu tego zagadnienia. Z kolei cechą wyróżniającą nowoczesne technologie w produkcji zwierzęcej jest coraz wyższy poziom automatyzacji zadań roboczych związanych np. z pozyskiwaniem mleka (roboty udojowe), kontrolą atmosfery/jakości powietrza w budynkach inwentarskich (roboty do czyszczenia rusztów), ale także stosowanie systemów pomagających zarządzać stadem zwierząt jako narzędzi umożliwiających podejmowanie

codziennych, ważnych dla wyników produkcyjnych, decyzji zw. z analizowaniem zachowań zwierząt w stadzie i ich aktualnej produktywności.

Świat nauki jest na etapie testowania nowoczesnych rozwiązań w rolnictwie, wykorzystujących sztuczną inteligencję do sterowania urządzeniami ułatwiającymi pozyskanie produktów pochodzenia roślinnego i zwierzęcego. Obecnie publikowane są prace naukowe w zakresie najlepszych algorytmów sztucznych sieci neuronowych, pozwalających poprawnie rozpoznać owoc i umożliwić jego zebranie poprzez zrobotyzowane ramię lub pojazd.

Automatyzacja obsługi klienta/pacjenta

W dzisiejszych czasach komputery świetnie radzą sobie z rozpoznawaniem mowy, czy pisma. Umożliwiło to stworzenie wielu użytecznych inteligentnych asystentów. Zadania, które kiedyś wymagały użycia klawiatury, możemy teraz coraz częściej wykonać tylko przy pomocy głosowych poleceń. Powszechniejsze staje się wykorzystanie technologii uczenia maszynowego w procesach obsługi klienta. Wszędzie tam, gdzie mamy do czynienia z dużą liczbą danych generowanych lub procesowanych w pewien powtarzalny sposób lub też, gdzie wymagane jest np. kategoryzowanie, mogą znaleźć zastosowanie rozwiązania oparte na sztucznej inteligencji. Właśnie w ten sposób działają m.in. najnowsze rozwiązania anty-spamowe, które bezustannie uczą się jak wygląda fałszywa lub niechciana korespondencja. Systemy te nie tylko analizują słowa kluczowe oraz swoistą specyfikę korespondencji (np. adresat, częstotliwość korespondencji), lecz również coraz częściej dokonują semantycznej oceny przesłanej treści.

Świat nauki, wykorzystując metody uczenia maszynowego, znacznie przyczynił się do rozwoju rozpoznawania mowy i automatycznej transkrypcji do wersji tekstowej. Obecnie największym wyzwaniem nie jest już dokonanie analizy samej mowy, a zrozumienie sensu wypowiedzi i przygotowanie poprawnej odpowiedzi. Innymi słowy cyfrowy asystent nie powinien reagować na słowa kluczowe jak np. „nieruchomości”, a rozumieć cały sens skierowanej do niego wypowiedzi użytkownika. Świat nauki i biznesu pracuje obecnie nad cyfrowymi asystentami wspierającymi konsumenta, obywatela czy profesjonalistę w zadanym obszarze (np. proces zakupowy). Brakuje jednak rozwiązań kompleksowo wspierających obywatela w procedowaniu spraw w instytucjach użyteczności publicznej.

Szczególnie istotne znaczenie może to mieć w przypadku lekarzy, pracujących w służbie zdrowia, którzy znaczną część czasu muszą przeznaczać na zapisywanie informacji. Na kartach chorego zapisywany jest wywiad z chorym, podstawowe informacje o jego stanie zdrowia, podsumowanie wyników badań specjalistycznych oraz wybrane wyniki ważne z punktu widzenia diagnozy. Do tego dochodzi obowiązek przygotowania recept i dokumentów dla NFZ. Inteligentny dyktafon, zapisujący informacje podawane przez lekarza w trakcie badania i przechowujący informacje o wcześniejszych diagnozach i zapisywanych lekach, mógłby bardzo skrócić czas przeznaczony na jednego pacjenta, bez skracania efektywnego czasu przeznaczanego na badania.

Dostępne są już algorytmy dostatecznie dobrze przekształcające mowę na tekst. Sprzęgnięcie tych algorytmów z systemem ekspertowym mogłoby pomóc lekarzowi w zapisywaniu w sposób usystematyzowany informacji, uzyskiwanych w trakcie każdego badania. System analizowałby informacje dyktowane przez lekarza i wpisywałby je w odpowiednie pola stosownego formularza, a także do bazy danych.

Innym popularnym problemem występującym w firmach i instytucjach, obsługujących dużą liczbę klientów, jest odpowiednie przekierowywanie korespondencji. Wciąż znaczna część pism kierowana jest do instytucji tradycyjną formą pocztową. To znacząco wpływa na opóźnienia w obsłudze wniosku czy też skargi. Nawet w przypadku korespondencji elektronicznej często następuje kierowanie pism na ogólny adres e-mail lub (w przypadków wielu urzędów) na jedną ogólną skrzynkę ePUAP. Obecnie istniejące systemy informatyczne potrafią jedynie przyspieszyć lub rozwiązać tylko wybrany element całego procesu obsługi korespondencji. Potrafią dokonać rozpoznania pisma i zamienić je na pismo tekstowe (tzw. technologia OCR), dokonać przekierowania na właściwą skrzynkę e-mail urzędnika lub pracownika biurowego na bazie słowa kluczowego w tytule. Wciąż jednak brak jest kompleksowych rozwiązań, które wykorzystując sztuczną inteligencję nie tylko rozpoznają tekst, ale również samodzielnie skierują korespondencję do właściwej osoby. Celowe jest zatem wprowadzenie inteligentnego dyspozytora korespondencji w urzędach i firmach, który pozwoliłby skrócić czas obsługi klienta, jak również podnieść poziom oferowanych usług.

Przyglądając się bliżej działaniu administracji publicznej, łatwo wysnuć wniosek, że obywatele, potrzebujący załatwić sprawę w urzędzie lub innych jednostkach publicznych, często nie wiedzą jak to zrobić: do jakiego pokoju/okienka urzędnika powinni się udać, jaki druk wypełnić, czy też co wpisać w wymagane pola. W konsekwencji obywatele spędzają długi czas w biurze informacji urzędu, wypełniając błędnie wnioski lub zadają dużą liczbę pytań urzędnikom. Już na etapie systemu kolejkowego (automatyczny kiosk wydający numery) oraz wyboru okienka obywatele często nie wiedzą jaki przycisk wybrać w automacie przydzielającym numerek. Wszystko to powoduje, że czas obsługi obywatela znacznie się wydłuża. Funkcjonujące obecnie w Polsce proste systemy obsługi obywateli działają na zasadzie formy tekstowej (chat bot w e-gov) lub uproszczonych systemów informacyjnych, w których obywatel musi dokonać wyboru. Nie obsługują one jednak naturalnego języka komunikacji, jakim jest mowa. Zastosowanie zaawansowanego asystenta petenta w instytucjach użyteczności publicznej ma szansę przyczynić się do skrócenia kolejek w urzędach oraz sprawniejszej obsługi petenta.

Analiza treści publikowanych w Internecie

Internet jest ważnym źródłem użytecznych informacji w ogromnej skali. Jednak od dłuższego czasu w Internecie pojawia się coraz więcej informacji mało wiarygodnych, co czasem jest wynikiem braku dbałości, ale coraz częściej niestety wynikiem celowych kampanii, wykorzystywanych do rozmyślnego rozpowszechniania fałszywych lub wprowadzających w błąd wiadomości. Sytuacja ta może stwarzać poważne zagrożenie dla zdrowia oraz finansowego bezpieczeństwa wielu osób, a zmasowane rozpowszechnianie fałszywych informacji może zaburzać procesy demokratyczne. Dlatego potrzeba narzędzi pozwalających na rozpoznawanie fałszywych treści. Prace nad takimi systemami prowadzi dla przykładu OpenAI, jednakże ze względu na możliwość wykorzystania takich narzędzi do generowania fałszywych informacji nie planuje udostępniać ich publicznie.

Blockchain

Powstanie gospodarki cyfrowej wydaje się w tym momencie nieuniknione, a jasną do tego przesłanką jest powstanie kryptowalut. Jednakże istniejące systemy blockchain nadal jeszcze mało wydajnie wspierają możliwość realizowania inteligentnych kontraktów, tzn. wykonywania umów między dwiema firmami w językach programowania z późniejszą gwarancją, że zostaną one zrealizowane. Możliwość efektywnego i bezpiecznego realizowania tego typu kontraktów wydaje się być konieczna do dalszego rozwoju gospodarki cyfrowej. Powstanie takiego systemu oznaczałoby dla firm oszczędności na kosztach prawnych oraz pozwoliłoby zminimalizować potencjalne spory.

Analiza warunków prowadzenia prac B&R w ICT

Kolejnym aspektem, na który zwracają uwagę eksperci, a który wciąż charakteryzuje polski B+R w ICT, jest długotrwałe prowadzenie badań koncepcyjnych i przecenianie znaczenia badań podstawowych w projektach technologicznych. Jest to zasadniczo cecha immanentna każdego ośrodka akademickiego, czy instytutu badawczego w Polsce. Przyzwyczajone są one bowiem do realizacji projektów trwających kilka lat, których oczekiwanym rezultatem jest raport, zawierający wyłącznie analizę. W opozycji do tego podejścia, biznes oczekuje od projektów, by szybko dochodziły do fazy poddającej się sprawdzeniu w działaniu (ang. *proof of concept*), czyli minimalnego działającego systemu. Objawem tego problemu może być m.in. wciąż rzadka współpraca nauki z biznesem w Polsce. Problem ten jest jeszcze silniej odczuwany w sektorze ICT, gdzie dojście do fazy *proof of concept* jest bardzo szybkie, a mimo to nawet w tej dziedzinie ośrodki akademickie prowadzą długotrwałe badania (nawet w przypadku, gdy dane do analizy są już dostępne, bądź są zbierane w sposób usystematyzowany).

W związku z powyższym aspekt dostępności danych, a także miar odniesienia (ang. *benchmarks*), jest stawiany przez ekspertów jako kolejne wyzwanie, które wymaga systematycznego rozwiązania. Dla wielu zastosowań takie dane nie są aktualnie dostępne, zaś nowo powstające firmy nie mają zasobów, by doprowadzić do ich udostępnienia w domenę publicznej lub takich danych w ogóle nie tworzą.

Oczywiście w niektórych dziedzinach zastosowań SI dostępność danych nie jest problemem, jak chociażby analiza zdjęć satelitarnych, gdzie dostępnych jest wiele różnorodnych repozytoriów danych. W takich przypadkach problemem wydaje się raczej brak uniwersalnych i łatwych do wdrożenia technik pozwalających na analizę tych danych. Stąd potrzeba stworzenia generycznych rozwiązań, na których możliwa będzie budowa bardziej zaawansowanych produktów.

Analiza potencjału B&R w dziedzinie sztucznej inteligencji

Potencjał B&R w dziedzinie, do której głównie odnosi się program INFOSTRATEG, został dobrze przedstawiony w następujących trzech dokumentach:

- aktualnie konsultowanej Polityce Rozwoju Sztucznej Inteligencji w Polsce na lata 2019 - 2027,
- raportu Badacze SI i Ich Publikacje, który został przygotowany przez OPI,
- Mapy polskiego SI przygotowanej przez fundację Digital Poland.

Dokumenty te wskazują na następujące aspekty:

- znaczną liczbę badaczy zajmujących się dziedziną sztucznej inteligencji – ponad 6,5 tysiąca badaczy publikujących w tej dziedzinie, z czego ponad 1/3 naukowców opublikowała już co najmniej 3 prace w ostatnich 6-ciu latach;
- niewielką liczbę firm, które prowadzą prace B&R w dziedzinie sztucznej inteligencji – tylko około 30 firm budujących SI.

Innymi słowy Polska posiada dość istotny potencjał badawczy w dziedzinie SI, jednakże potencjał ten nie przekłada się na działalność B&R. W szczególności Polityka Rozwoju Sztucznej Inteligencji wskazuje, że aktywacja potencjału badawczego może wymagać dodatkowych nakładów finansowych rzędu około 8 mld PLN. Ze względu na te uwarunkowania, INFOSTRATEG będzie stanowił istotny wkład w tę aktywację. Szersza analiza wyzwań oraz przeszkód w zakresie efektywnego wykorzystania sztucznej inteligencji znajduje się w Polityce Rozwoju Sztucznej Inteligencji.

Analiza zgodności z dokumentami strategicznymi

Tematyka badawcza przedstawiona w programie INFOSTRATEG wpisuje się w następujące dokumenty strategiczne:

- Odpowiada jednemu z trzech obszarów priorytetowych „Strategii Europa 2020: Wzrost inteligentny”, czyli rozwój gospodarki opartej na wiedzy i innowacjach.

- Zakres tematyczny programu jest zgodny z „Długookresową Strategią Rozwoju Kraju (DSKR) – Polska 2030. Trzecia fala nowoczesności”, który podkreśla wagę wykorzystania technik cyfrowych.
- „Wzmocnienie cyfrowego rozwoju kraju” jest jednym z obszarów, które wpływają na osiągnięcie celu Strategii na rzecz Odpowiedzialnego Rozwoju – automatyzacja, robotyzacja oraz informatyzacja procesów gospodarczych.
- Założenia i tematyka programu INFOSTRATEG wpisują się w drugi cel „Strategii Innowacyjności i Efektywności Gospodarki: Stymulowanie innowacyjności poprzez wzrost efektywności wiedzy i pracy”.
- Program INFOSTRATEG wpisuje się zarówno w cel strategiczny Krajowego Programu Badań (wykorzystanie nauki dla podniesienia poziomu cywilizacyjnego Polski, m.in. poprzez pełniejsze wdrożenie jej wyników w edukacji, gospodarce i kulturze) jak i w priorytetowe kierunki badań naukowych i prac rozwojowych ujęte w Krajowym Programie Badań.
- Wszystkie tematy wskazane do realizacji w Programie wpisują się w dziesiątą Krajową Inteligentną Specjalizację.
- Temat T11 mieści się w Krajowej Ramie Polityki Cyberbezpieczeństwa Rzeczypospolitej Polskiej (KRPC RP) na lata 2017-2022.
- Program INFOSTRATEG jest w bardzo dobrej synergii z aktualnie konsultowaną Polityką Rozwoju Sztucznej Inteligencji w Polsce na lata 2019-2027.

3. Cel główny i cele szczegółowe

Z przeprowadzonej w poprzednim rozdziale analizy sytuacji w obszarach nauki i gospodarki wynika, że nowoczesne metody uczenia maszynowego będą coraz powszechniej wykorzystywane. Aby umożliwić polskim firmom odnalezienia się w tych nowych warunkach biznesowych, B+R w Polsce powinny nadążyć za tym trendem. Oczywiście rozwijanie nowoczesnych technologii wymaga wskazania ich konkretnego zastosowania – możliwa jest wówczas dokładna ocena rozwiązania i porównanie go z innymi dostępnymi na rynku. W Programie INFOSTRATEG priorytetem jest wskazanie właśnie takich zastosowań, bądź też wyłonienie ich w ramach konkursów zamawianych. Prowadzi to tym samym do sformułowania celu głównego Programu.

Celem głównym programu jest rozwój polskiego potencjału SI poprzez opracowanie rozwiązań wykorzystujących sztuczną inteligencję i blockchain, mających bezpośrednie zastosowanie w praktyce.

Nieodłącznym celem tego Programu jest powstanie dla każdego ze wskazanych zastosowań środowiska testowego pozwalającego na dobrą ocenę i porównanie między sobą różnych rozwiązań. Jedną z głównych przeszkód w badaniach nad uczeniem maszynowym w Polsce jest, jak już wspomniano, brak dostępnych danych. Program INFOSTRATEG ma się przyczynić do zmiany tej sytuacji.

Cel szczegółowy I: Utworzenie zbiorów danych testowych i stworzenie na ich bazie standardów wyboru najlepszych rozwiązań.

W szczególności cel ten odnosi się do tematów: T1, T2, T3, T4, T6, T7, T8, T10, T11 oraz T12. W przypadku realizacji tych tematów zakładamy, że zbiory danych testowych powstaną poprzez połączenie ze sobą danych pochodzących od wszystkich podmiotów realizujących projekty. Zebranie dobrej jakości danych oraz stworzenie na ich podstawie standardów oceny ma umożliwić obiektywną ocenę prezentowanych rozwiązań. Natomiast stworzenie rozwiązań konkurencyjnych na światowym poziomie wymaga wbudowania w realizację Programu konkurencji wewnętrznej, tzn. dla każdego zastosowania należy wyłonić w konkursie kilka zespołów, które będą ze sobą konkurowały w danym temacie. Taka konkurencja przyczyni się do powstawania rozwiązań o bardzo dobrej jakości, a także do wytworzenia istotnych rezultatów badawczych, co bezpośrednio prowadzi do kolejnego celu szczegółowego tego programu.

Cel szczegółowy II: Rozwijanie polskiego potencjału badawczego w sztucznej inteligencji.

Cel ten odnosi się do wszystkich tematów: T1-T12. Oczekuje się, że projekty będą realizowane przy pomocy metodologii zwinnych (ang. *agile*). Ma to na celu zmianę podejścia do prowadzenia B+R w ośrodkach akademickich poprzez stymulowanie do maksymalnie szybkiego tworzenia przez nie rozwiązań typu *proof of concept*. Oczekuje się także, że projekty będą realizowane zgodnie z polityką *Cloud First*, tzn. zakup sprzętu będzie mógł być realizowany tylko i wyłącznie, jeśli nie jest możliwa realizacja z użyciem usługi przetwarzania w chmurze.

Warto zaznaczyć, że istotnym aspektem przy tworzeniu kolejnych, bardziej zaawansowanych rozwiązań, jest dostępność bibliotek i rozwiązań w postaci oprogramowania typu *open source*. Program INFOSTRATEG powinien przyczynić się do powstania otwartego oprogramowania. Stworzenie szerokiej bazy rozwiązań (także dostępnych w formule otwartej) znacząco ułatwi i zmniejszy koszt tworzenia nowych rozwiązań dla powiązanych problemów. W zastosowaniach informatyki przełomowe techniki, które z sukcesem udało się zastosować dla wybranego problemu zwykle znajdują dużo szersze zastosowanie. Często kluczowa jest właśnie dostępność pewnej bazy oprogramowania, którą można by było odpowiednio modyfikować do kolejnych zastosowań. W szczególności powstanie pewnej liczby generycznych rozwiązań dla typowych problemów pozytywnie przyczyni się do realizacji następującego celu szczegółowego tego programu.

Cel szczegółowy III: Znaczące zwiększenie aktywności rynkowej polskich zespołów informatycznych.

Cel ten odnosi się do wszystkich tematów: T1-T12.

Stworzenie podstawowej bazy, oprogramowania, know-how i metodologii, powiązanych z uczeniem maszynowym, spowoduje znaczące obniżenie kosztu tworzenia nowych rozwiązań. Zachęci to nowe zespoły do zapoznania się z tym tematem i umożliwi łatwiejsze rozpoczęcie prac badawczych, a także wdrożeniowych – doprowadzi zatem do zwiększenia aktywności rynkowej polskich zespołów informatycznych oraz firm.

Cel szczegółowy IV: Zastosowanie sieci neuronowych w robotyce i automatyzacji.

W szczególności cel ten odnosi się do tematu: T5.

Automatyzacja z wykorzystaniem sztucznej inteligencji oraz rozwiązań ICT istotnie wpłynie na wzrost efektywności i jakości produkcji, obniżenie kosztów wytwarzania oraz pośrednio poziom życia. Efektem projektów mają być przetestowane prototypy inteligentnych maszyn, które mają potencjał do wdrożenia na skalę przemysłową. W ten sposób na rynek polski trafią pionierskie rozwiązania wykorzystujące SI, będące praktyczną odpowiedzią na wyzwania istniejące w tym momencie w rolnictwie.

Cel szczegółowy V: Stworzenie narzędzi opartych o technologię blockchain przyspieszających rozwój gospodarki cyfrowej.

Cel ten odnosi się do tematu: T9.

Opracowane w ramach projektów narzędzia oparte na technologii blockchain będą istotnym akceleratorem rozwoju gospodarki cyfrowej. Zastosowanie takich systemów w obszarze eKontraktów przyniosłoby firmom znaczne korzyści, w tym zwiększenie bezpieczeństwa, oszczędności na kosztach obsługi kontraktu i usług prawnych, a także pozwoliłoby zminimalizować potencjalne spory.

Cel szczegółowy VI: Stworzenie rozwiązań opartych o uczenie maszynowe podnoszących jakość produktów/ usług i efektywność procesów.

W szczególności cel ten odnosi się do tematów: T1, T2, T3, T4, T6, T7, T8, T9, T10, T11 oraz T12.

Istnieją obszary, gdzie maszyna szybciej i lepiej niż człowiek wykona zadanie. Zakłada się, że w wyniku projektów wykorzystujących potencjał sztucznej inteligencji, opracowane zostaną praktyczne rozwiązania oparte o uczenie maszynowe, gdzie czasochłonne, zasobochłonne zadania mogą być szybciej i bezbłędnie realizowane przez komputer, prowadząc do wzrostu jakości usług, efektywności procesów oraz wsparcia analizy danych i procesu podejmowania decyzji.

4. Zakres tematyczny Programu

W rozdziale tym podano tematy, które możliwe są do realizacji w ramach Programu. Tematy te zostały wyselekcjonowane jako posiadające duży potencjał wdrożeniowym, a także charakteryzujące się koniecznością stworzenia zaawansowanych rozwiązań informatycznych, tzn. wykorzystujących algorytmy uczenia maszynowego, czy techniki blockchain. Podane przy tematach budżety są orientacyjne. Ostateczne budżety zostaną określone przez Komitet Sterujący. Zaproponowane tematy badawcze spełniają następujące wymagania:

- jeżeli do prowadzenia przewidzianych badań potrzebne są dane testowe, należy precyzyjnie określić w jaki sposób mogą one zostać zdobyte, bądź wytworzone w ramach realizacji projektów;
- realizacja tematów powinna wymagać użycia nowoczesnych narzędzi uczenia maszynowego, bądź dalszego rozwoju technologii blockchain;
- w tematach ma dominować aspekt stworzenia oprogramowania, choć komponent sprzętowy może w nich występować;
- dla zaproponowanych tematów powinien istnieć jasno określony, ilościowy sposób porównywania między sobą różnych rozwiązań.

Zakres tematyczny tego programu może być rozszerzony o tematy projektów zamawianych, spełniające warunki wymienione powyżej i wyselekcjonowane w sposób określony w rozdziale 5. W szczególności tematy mogą dotyczyć kluczowych sektorów takich jak: przemysł, logistyka, energetyka czy transport. Warto zaznaczyć, że mimo iż program stawia na rozwój oprogramowania, w wielu przypadkach realizacja tematów wymaga włączenia ekspertów dziedzinowych do zespołów badawczych.

T1) Inteligentny system przetwarzania mowy dla lekarzy

Szacowany Budżet: 30 mln PLN.

Sytuacja obecna. Lekarze pierwszego kontaktu połowę czasu przeznaczonego na wizytę spędzają na wypełnianiu dokumentacji medycznej, wypisywaniu recept i wprowadzaniu danych. Przyspieszenie i ułatwienie tego procesu przyczyni się bezpośrednio do poprawienia jakości obsługi pacjentów. Prace nad tego typu systemami prowadzone są już na świecie (np. suki.ai), jednakże systemy te przeznaczone są na razie dla języka angielskiego i nie są dostosowane do wymagań polskiego systemu zdrowia. Przeprowadzony Hakaton w trakcie forum eZdrowia w 2019 wykazał, że prywatne polskie instytucje medyczne są bardzo zainteresowane takim rozwiązaniem.

Cel projektu. Celem projektów będzie opracowanie narzędzi pozwalających na przetwarzanie mowy (frazy wypowiedziane przez lekarza i ewentualnie pacjenta) i wydobywanie z tekstu istotnych informacji, które następnie byłyby klasyfikowane semantycznie i wprowadzane do wcześniej zbudowanego formularza i bazy danych. Formularze powinny być elastyczne, tak aby mogły być

modyfikowane wraz z rozwojem narzędzi medycznych oraz dostosowywane do specjalności lekarza, który z nich korzysta. Na podstawie zebranej wiedzy system mógłby też automatycznie generować dokumenty: recepty, zwolnienia, skierowania do dalszych badań diagnostycznych, itp.

Produkty projektu.

- Zbiór danych testowych zawierających zanonimizowane wywiady lekarskie z odpowiednio wypełnionymi formularzami.
- Benchmarki pozwalające określić jakość rozwiązań wymaganą do wdrożenia.
- Przetestowane rozwiązanie.

Korzyści.

- Skrócenie kolejek do lekarza, z uwagi na krótszy czas spędzony przez pacjenta w gabinecie lekarskim. Obecnie Polacy przeciętnie czekają 3,4 miesiąca na wizytę do lekarza.
- Część czasu jaki lekarza spędza na przepisywaniu danych do komputera będzie mógł spędzić na pogłębionej diagnozie i rozmowie z pacjentem podnosząc jakość leczenia i diagnozy.

T2) Rozpoznawanie obrazów medycznych

Szacowany budżet: 30 mln PLN.

Sytuacja obecna. Co roku na raka prostaty zapada w Polsce ponad 15 tysięcy mężczyzn, na raka płuc 15 tysięcy mężczyzn i 7 tysięcy kobiet, a na raka piersi około 15 tysięcy kobiet. Wiele z tych osób umiera, co jest związane z późną rozpoznawalnością tych chorób, a wynika pośrednio z braków kadrowych. Służba zdrowia zmagą się z ogromnym niedoborem kadr w wielu specjalnościach, niedobór ten jest szczególnie dotkliwy wśród ekspertów interpretujących dane obrazowe (histopatologów, specjalistów opisujących zdjęcia rentgenowskie, specjalistów opisujących dane pochodzące z bardziej zaawansowanych urządzeń t.j. tomografy komputerowe czy rezonans magnetyczny. Na skutek tego czas oczekiwania na wyniki jest bardzo długi, co opóźnia rozpoczęcie leczenia, zmniejszając szanse na całkowite wyleczenie, a także powoduje stres u pacjentów, zniechęcenie podejmowaniem dalszego leczenia oraz wpływa na ogólne obniżenie u społeczeństwa zainteresowania badaniami.

Aktualnie istniejące systemy rozpoznawania obrazów nie są jeszcze gotowe do użytku w praktyce medycznej ze względu na tzw. problem braku wyjaśnialności (ang. explainability), czyli podejmowania przez te systemy decyzji na podstawie nieadekwatnych przesłanek. Ministerstwo Zdrowia oraz Ministerstwo Nauki i Szkolnictwa Wyższego wyraziło wstępne zainteresowanie tym tematem w ramach prac przygotowawczych programu INFOSTRATEG.

Cel projektu. Celem projektów realizujących ten temat będzie opracowanie uniwersalnych narzędzi do analizy i opisu różnego typu obrazów medycznych. W szczególności stworzone techniki powinny:

1. dawać możliwość wyszukiwania i oznaczania różnych obszarów o różnej charakterystyce na zdjęciach;
2. rozpoznawać czy oznaczone obszary mogą być związane z jakąś chorobą;
3. dostarczać możliwe informacje analityczne dla badanych zdjęć i wyróżnionych obszarów;
4. wyjaśniać powody przedstawionej diagnozy, tzn. opisywać występujące patologiczne obszary na zdjęciu i wyjaśniać ewentualne przyczyny, dlaczego zostały tak zakwalifikowane.

Produkty projektu.

- Zbiory zanonimizowanych obrazów medycznych różnego typu (rentgenowskie, histopatologiczne, z tomografów komputerowych, itd.) z odpowiednimi adnotacjami i oznaczeniami. Zebrane zdjęcia powinny demonstrować różne scenariusze życia, co pozwoliłoby na zwiększenie uniwersalności generowanych opisów.
- Benchmarki pozwalające określić jakość rozwiązań wymaganych do wdrożenia.
- Zbudowane przy pomocy uczenia maszynowego algorytmy wspierające pracę lekarzy, oznaczające patologicznie zmienione fragmenty obrazów i opisujące rodzaj zmian.

Korzyści.

- Poprawa jakości diagnostyki obrazowej.
- Skrócenie czasu oczekiwania na diagnozę.
- Umożliwienie wcześniejszego rozpoczęcia leczenia często występujących postaci raka.
- Zmniejszenie śmiertelności w przypadku często występujących i późno wykrywanych postaci raka.
- Skrócenie kolejek do lekarzy specjalistów.

T3) Diagnostyka wrodzonych wad metabolizmu w badaniach przesiewowych noworodków

Szacowany budżet: 20 mln PLN.

Sytuacja obecna. Populacyjne badania przesiewowe noworodków w kierunku chorób rzadkich prowadzone w Polsce przez Instytut Matki i Dziecka pozwalają na wykrycie choroby w fazie przedobjawowej. Badania te obejmują następujące choroby: fenyloketonurię, wrodzoną niedoczynność tarczycy, mukowiscydozę, głuchotę oraz od niedawna rzadkie wrodzone wady metabolizmu. Badania polegają na poddawaniu próbek krwi lub moczu pacjenta analizie przy pomocy chromatografii gazowej i spektrometrii mas. W wyniku pomiaru powstają profile, których

interpretacja wymaga czasochłonnych ocen przez wykształconych specjalistów, co ogranicza masowość i różnorodność wykonywanych badań, prowadząc do różnego rodzaju opóźnień.

Jest to temat, który leży w bezpośrednim zainteresowaniu Instytutu Matki i Dziecka. Także Ministerstwo Zdrowia i Ministerstwo Nauki i Szkolnictwa Wyższego wyraziło wsparcie dla tematów, związanych z wsparciem diagnostyki w ramach prac przygotowawczych programu INFOSTRATEG.

Cel projektu. Celem projektów realizujących ten temat byłoby zbudowanie i przetestowanie algorytmu analizy plików powstających w wyniku pomiarów na urządzeniach chromatografii gazowej i spektrometrii mas. Algorytm docelowo byłby używany do badań przesiewowych w Instytucie Matki i Dziecka, ale mógłby z powodzeniem służyć innym ośrodkom dysponującym odpowiednią aparaturą.

Produkty projektu.

- Zanonimizowane zbiory danych pomiarowych z badań prowadzonych na urządzeniach chromatografii gazowej i spektrometrii mas, które zawierają wyniki ręcznej analizy.
- Benchmarki pozwalające określić jakość rozwiązań.
- Efektywne metody detekcji i adnotacji sygnału.
- Przetestowany system informatyczny wspomagający diagnostykę wrodzonych wad metabolizmu u noworodków, wykorzystujący techniki uczenia maszynowego do w pełni automatycznej interpretacji widm spektrometrycznych.

Korzyści.

- Poprawa diagnostyki wrodzonych wad metabolizmu u noworodków.
- Skrócenie czasu oczekiwania na wynik i uniknięcie zagrożeń wynikających z opóźnień (w niektórych chorobach skuteczne leczenie musi rozpocząć się w ciągu kilku dni po urodzeniu).
- Możliwość rozszerzenia zakresu prowadzonych badań przesiewowych.

T4) Różne scenariusze rozpoznawania zdjęć satelitarnych i lotniczych

Szacowany budżet: 40 mln PLN.

Sytuacja obecna. Zdjęcia satelitarne potencjalnie dają możliwość automatycznego wywnioskowania różnego rodzaju informacji geolokalizacyjnej. Jednakże brakuje narzędzi, które w łatwy sposób pozwalałyby na rozpoznawanie nowych typów obiektów, czy obiektów o różnym charakterze. Stworzenie uniwersalnych narzędzi analizy zdjęć satelitarnych i lotniczych pozwoli na ich wykorzystanie w różnych obszarach działalności, np. rolnictwie. Warto zaznaczyć, że na świecie od niedawna powstają start-upy, które mają na celu stworzenie tego typu narzędzi np. SpaceKnow.

Jednakże wydaje się, że możliwości tego typu rozwiązań dalekie są od osiągnięcia pełnej uniwersalności.

W ramach prac przygotowawczych do programu INFOSTRATEG zainteresowanie tym tematem wyrazili następujący interesariusze: Ministerstwo Infrastruktury (dawniej: Ministerstwo Infrastruktury i Budownictwa), Ministerstwo Gospodarki Morskiej i Żeglugi Śródlądowej, Ministerstwo Klimatu (dawniej: Środowiska) oraz Kancelaria Prezesa Rady Ministrów.

Cel projektu. Celem Programu jest stworzenie uniwersalnych technik pozwalających na rozpoznawanie wielorakich obiektów na zdjęciach satelitarnych i lotniczych. W szczególności stworzone techniki powinny:

- dawać możliwość wyszukiwania nowych rodzajów obiektów, np. poprzez podanie tylko kilku (czy nawet jednego) przykładów takiego obiektu, np. samochodu konkretnej marki i koloru, czy pola o zbliżonym poziomie agrotechniki;
- rozpoznawać klasy użytkowania terenu i zmiany tego użytkowania w czasie;
- rozpoznawać obiekty o różnych rozmiarach linearnych, czy o różnej powierzchni, np. drogi, rzeki, pola, czy lasy;
- dostarczać podstawowe informacje analityczne, np. ilość, powierzchnia, długość, czy też oceniać zmiany tych wartości;
- dostarczać bardziej zaawansowane informacje analityczne dot., np. zróżnicowania warunków środowiskowych w regionie, rozpoznawania ognisk i dynamiki rozprzestrzeniania się chorób roślin, interakcji między zmianami użytkowania terenu a charakterem produkcji rolniczej, itp.;
- pozwalać na monitorowanie stanu zachowania nieruchomych zabytków dziedzictwa, poprzez np. porównywanie zdjęć z różnych okresów czasowych;
- ewentualnie umożliwiać monitoring przestrzegania zasad zrównoważonej produkcji i efektów adaptacji rolnictwa do zmian klimatu oraz oceniać wpływ praktyk rolniczych na środowisko.

Produkty projektu.

- Szeroki zbiór danych testowych, które demonstrowałyby różne scenariusze użycia.
- Dla każdego scenariusza użycia benchmark jakościowy.
- Przetestowane uniwersalne algorytmy adnotacji zdjęć satelitarnych.

Korzyści.

- Aktualizacja danych systemów geodezyjnych i kartograficznych z wykorzystaniem zdjęć satelitarnych i zdjęć z dronów, obserwacja pełzania i osuwania się gruntów.
- Teledetekcja związana ze środowiskiem wodnym, śledzenie wędrówek ryb, monitorowanie żeglugi.
- Monitorowanie zanieczyszczeń.
- Budowa inteligentnych usług informacyjnych w oparciu o zasoby danych satelitarnych.

T5) Inteligentna maszyna do zbierania jabłek

Szacowany budżet: 100 mln PLN.

Sytuacja obecna. W Polsce w 2018 roku wyprodukowano 4.8 miliona ton jabłek. Z tego połowa pochodziła z nowoczesnych, wysoko dochodowych sadów w większości o powierzchni 15 – 20 hektarów, uzyskujących plony 40 – 50 ton/ha i produkujących głównie dobrej jakości jabłka deserowe, atrakcyjnych rynkowo odmian. Właściciele tych sadów (jest ich około 3000) będą (są) zainteresowani zakupem robotów do automatycznego zbioru. W kilku miejscach na świecie prowadzone są prace nad zbudowaniem robotów do zbioru jabłek (prace jeszcze wciąż trwają). Dodatkowo, sposób budowania sadów w różnych krajach jest odmienny i maszyny wyprodukowane na potrzeby sadowników w jednym kraju mogą być nieodpowiednie dla sadowników z innego kraju.

Polska w niedalekiej przyszłości zmagać się będzie z kryzysem demograficznym, przez co w rolnictwie najprawdopodobniej zabraknie pracowników. Automatyzacja zbierania owoców jest swego rodzaju odpowiedzią na spodziewany problem.

Cel projektu. Projekty realizowane w tym temacie będą miały na celu opracowanie inteligentnej maszyny, umożliwiającej zbieranie jabłek w sposób zautomatyzowany, dzięki m.in. wykorzystaniu sztucznej inteligencji. Inteligentna maszyna powinna opierać się na oprogramowaniu, które pozwoli rozpoznać owoc, a następnie dzięki wykorzystaniu np. robotycznego ramienia, dokonać jego zbioru. Przewiduje się wykorzystanie komercyjnie dostępnych maszyn czy ramion robotów, zaś głównym celem projektów powinno być stworzenie prototypu maszyny wraz z oprogramowaniem.

Produkty projektu.

- Środowisko testowe pozwalające określić jakość rozwiązań wymaganych do wdrożenia.
- Prototyp maszyny wraz z oprogramowaniem (przetestowanej w warunkach polowych) i dokumentacją techniczną umożliwiającą jej produkcję na skalę przemysłową.

Korzyści.

- Znaczące obniżenie kosztu zbierania dobrej jakości jabłek deserowych i zwiększenie dochodów polskich sadowników.

- Inteligentne maszyny do zbioru jabłek po przetestowaniu w Polsce w warunkach polowych będą mogły być sprzedawane do innych krajów produkujących jabłka i owoce o podobnych parametrach (Węgry, Ukraina, Białoruś, itp.).
- Opracowane technologie będą mogły zostać użyte do produkcji robotów zbierających inne owoce (w szczególności owoce miękkie).

T6) Scenariusze selektywnej ochrony roślin

Szacowany budżet: 40 mln PLN.

Sytuacja obecna. Obecnie, powszechną i jedyną techniką aplikowania przez rolników środków chemicznych i nawozów jest stosowanie zalecanych dawek równomiernie na całej plantacji. To samo dotyczy nawadniania. Według danych statystycznych w Polsce jest około 27 tysięcy gospodarstw rolniczych średnich (50 – 200 ha), które powszechnie stosują wybrane metody rolnictwa precyzyjnego oraz około 5 tysięcy gospodarstw dużych, o powierzchni powyżej 200 ha, które stosują wszystkie dostępne produkty rolnictwa precyzyjnego. Są one potencjalnie zainteresowane selektywnym nawożeniem i ochroną roślin.

Cel projektów. Projekt będzie miał na celu opracowanie algorytmów rozpoznających, na podstawie obrazu z kamer i czujników zainstalowanych na ciągnikach, wystąpienie niepożądanych zjawisk na plantacjach t.j. choroby roślin i inwazje (wysyp) szkodliwych owadów, czy też nieprawidłowy rozwój wynikający z braku wilgoci lub pożądaných substancji w glebie. Uzyskane w ten sposób informacje, uzupełnione o inne dostępne dane, byłyby stosowane w komputerach pokładowych ciągników (i innych agro-komputerach) do sterowania zespołami wykonawczymi maszyn rolniczych m.in. do dozowania środków ochrony roślin i nawozów. Zbudowany system będzie miał zastosowanie co najmniej przy zwalczaniu różnych szkodników roślin poprzez ustalanie optymalnego dawkowania środków chemicznych.

Produkty projektu.

- Zbudowane przy pomocy uczenia maszynowego algorytmy, rozpoznające niepożądane zjawiska w określonych fragmentach plantacji i określające ich przyczyny (np. zniszczone pączki kwiatów rzepaku w wyniku ataku określonego szkodnika). Stworzone urządzenia (algorytmy) powinny uwzględniać specyfikę polskiego rolnictwa (np. duża koncentracja produkcji rzepaku).
- Algorytmy mapujące zebrane informacje do systemów sterujących GPS dla urządzeń aplikujących/dozujących środki chemiczne/nawozy.
- Różne zbiory danych testowych – otagowanych filmów stworzonych przez kamery zainstalowane na maszynach rolniczych.
- Benchmarki pozwalające określić jakość rozwiązań wymaganych do wdrożenia.

Korzyści.

- Zmniejszenie kosztów nawozów i środków ochrony roślin.
- Zmniejszenie zanieczyszczenia gleby i wód gruntowych szkodliwymi środkami chemicznymi oraz nadmiarem związków azotowych.
- Poprawa jakości zbiorów (produkcja żywności ekologicznej).
- Zwiększenie plonów.
- Pojawienie się możliwości eksportu produktów i technologii (po przetestowaniu na polskim rynku).

T7) Inteligentny dyspozytor korespondencji

Szacowany budżet: 20 mln PLN.

Sytuacja obecna. W dzisiejszych czasach coraz powszechniejszy staje się kontakt z organami administracji państwowej przez emaila lub inne formy komunikacji elektronicznej. Taki sposób kontaktu potencjalnie umożliwia przetwarzanie tych dokumentów w sposób automatyczny. Prace nad rozwijaniem odpowiednich technologii ICT trwają (gł. dla zastosowań w firmach prywatnych, np. Alphamoon), jednakże systemy te w tym momencie dostarczają bardzo ograniczoną funkcjonalność, tzn. są w stanie klasyfikować korespondencje, ale nie wspierają procesu przygotowywania odpowiedzi.

Na konieczność rozwijania technologii ICT w ramach administracji publicznej wskazało Ministerstwo Rodziny, Pracy i Polityki Społecznej.

Cel projektu. Celem projektów w ramach tego tematu będzie opracowanie inteligentnego dyspozytora korespondencji wspierającego proces obsługi obywatela lub klienta w języku polskim. Inteligentny dyspozytor korespondencji powinien być oprogramowaniem możliwym do zainstalowania na komputerze/serwerze, który na wejściu (input) otrzyma korespondencję, a na wyjściu:

- wskaże typ sprawy jakiej dotyczy korespondencja, dokona jej właściwej kategoryzacji oraz skieruje ją do właściwego urzędnika lub pracownika korporacji,
- na podstawie treści listu i historycznych danych zaproponuje szablon odpowiedzi, bądź znajdzie odpowiedzi na podobne zapytania,
- wywnioskuje numer sprawy i dane interesanta, jeżeli dane te nie są podane wprost,
- dostarczy skrótowej informacji o temacie korespondencji.

Produkty projektu.

- Zbiór danych testowych, zawierających skategoryzowaną korespondencję, oraz sugerowane odpowiedzi.
- Benchmark określający jakość działania algorytmów.
- Przetestowany system, który realizuje wyżej opisane funkcjonalności.

Korzyści.

- Krótszy proces obsługi klienta.
- Szybsze załatwienie jego sprawy w urzędzie lub firmie.

T8) Asystent petenta w instytucjach użyteczności publicznej

Szacowany budżet: 30 mln PLN.

Sytuacja obecna. Coraz powszechniejsze w użyciu stają się systemy asystentów głosowych, np. Siri, Asystent Google itp. Systemy te są w stanie w przystępny dla wszystkich sposób udostępniać potrzebne informacje. Niektóre z nich także bardzo dobrze radzą sobie z obsługą konwersacji w języku polskim. Dostarczanie informacji w sposób głosowy w instytucjach administracji/ użyteczności publicznej jest więc naturalnym dalszym krokiem w rozwoju takich narzędzi.

Na konieczność rozwijania takich narzędzi wskazało Ministerstwo Rodziny, Pracy i Polityki Społecznej, a także Ministerstwo Zdrowia, które zaproponowało stworzenie systemu dostarczającego informacji o procedurach medycznych.

Cel projektu. Projekty realizujące ten temat będą miały na celu opracowanie uniwersalnego cyfrowego asystenta wspierającego obywatela posługującego się językiem polskim w kontaktach z instytucjami użyteczności publicznej. Cyfrowy asystent mógłby być instalowany w urzędzie na dedykowanym sprzęcie (np. tablecie), jak również na smartfonie obywatela, co przełożyłoby się na lepsze przygotowanie obywatela do załatwienia sprawy w urzędzie.

Produkty projektu.

- Zbiór testowy zawierający przykładowe konwersacje między petentem a urzędem.
- Benchmarki pozwalające określić jakość rozwiązań wymaganych do wdrożenia.
- Przetestowane rozwiązanie.

Korzyści.

- Obniżenie kosztów obsługi petentów w instytucjach publicznych.
- Przyspieszenie obsługi petentów w instytucjach publicznych.

T9) Wydajny i skalowalny blockchain konsorcyjny dla inteligentnych kontraktów

Szacowany budżet: 40 mln PLN.

Obecna sytuacja. Przez *blockchain konsorcyjny* rozumie się system, w którym z góry określona grupa N serwerów ("członków konsorcjum") utrzymuje wspólną, publicznie dostępną tablicę z danymi. Danych tych nie można kasować. Prawo do publikacji danych (zwanymi "*transakcjami*") na tablicy posiadają wybrani użytkownicy systemu. Przez funkcję *wykonywania inteligentnych kontraktów* rozumie się możliwość umieszczania i egzekucji inteligentnych kontraktów zapisanych w jednym z popularnych języków programowania (np. *Solidity* lub *Go*).

System taki powinien zawierać metodę dodawania i usuwania nowych członków konsorcjum, a także dawać możliwość zmiany reguł publikowania informacji na tablicy oraz uaktualniania wersji systemu (w tym modyfikacji składni i semantyki języka inteligentnych kontraktów). Powinno się to odbywać w sposób zdecentralizowany (poprzez mechanizm głosowania).

System powinien być odporny na dowolne nieuczciwe zachowania określonej liczby T użytkowników (tzw.: "korupcje bizantyjskie"). Wraz ze specyfikacją systemu należy przedstawić dowód jego bezpieczeństwa. W analizie bezpieczeństwa wolno korzystać ze standardowych założeń kryptograficznych (bezpieczeństwo schematu podpisu, założenie o losowej wyroczni, itp.).

Cel projektu. Projekty realizujące ten temat miałyby za zadanie opracować działający system blockchainu konsorcyjnego, wspierającego wykonywanie inteligentnych kontraktów. Jakość i efektywność takiego systemu byłaby mierzona przy pomocy następujących parametrów:

- I. *Finality*, tj. czasu potrzebnego do uzyskania pewności, że dana transakcja nie może zostać usunięta. Czas ten może być uzależniony od stopnia pewności nieusuwalności transakcji (np.: "po 1 minucie pewność co najmniej 99%");
- II. *Throughput*, tj. liczby transakcji na sekundę mierzonej osobno w bajtach oraz w pojedynczych transakcjach;
- III. Złożoności komunikacyjnej protokołu (mierzona osobno liczbą wiadomości oraz ich rozmiarem);
- IV. Maksymalnej liczby tolerowanych nieuczciwych użytkowników T ;
- V. Złożoności obliczeniowej (obliczenie nie powinno być kosztem dominującym w porównaniu z komunikacją).

Parametry wydajnościowe (tzn.: I, II, III, i V) można oceniać osobno w warunkach: "optymistycznym" i "pesymistycznym". Przez pierwszy należy rozumieć scenariusz, w którym całość (bądź znaczna

część) konsorcjum zachowuje się uczciwie. Przez drugi rozumiemy sytuację, w której nieuczciwi członkowie konsorcjum złośliwie starają się maksymalnie wydłużyć czas działania protokołu.

Wydajność protokołu można mierzyć w warunkach rzeczywistych, tzn. w sieci Internet w sytuacji, w której serwery tworzące konsorcjum znajdują się w znacznej odległości (w różnych krajach, na różnych kontynentach). Pomiary wydajności będą wykonane dla (a) małej liczby ($N = 5$), (b) średniej ich liczby ($N=20$), oraz (c) dużej liczby serwerów ($N=100$).

Dopuszcza się użycie technik optymalizacyjnych, w których blockchain nie będzie informowany o wszystkich transakcjach w systemie, o ile dzieje się to za zgodą wszystkich zainteresowanych stron danej transakcji (tzw. techniki "off-chain"), oraz technik polegających na podziale konsorcjum na mniejsze grupy (tzw. "sharding").

Produkty projektu.

- System pozwalający na zawieranie kontraktów między firmami w sposób elektroniczny.
- Benchmarki pozwalające określić jakość rozwiązań wymaganych do wdrożenia.

Korzyści.

- Stosowanie eKontraktów oznaczałoby dla firm oszczędności na kosztach prawnych oraz pozwoliłoby zminimalizować potencjalne spory.
- Zagadnienie to jest wskazywane jako kolejne bardzo obiecujące zastosowanie technologii Blockchain.

T10) Rozpoznawanie na filmie zachowań ludzkich

Szacowany budżet: 40 mln PLN.

Sytuacja obecna. Systemy uczenia maszynowego wykorzystujące sieci neuronowe odnoszą spektakularne sukcesy w rozpoznawaniu obiektów na zdjęciach. Uważa się, że dostarczają one rezultatów dużo lepszej jakości niż człowiek, tzn. popełniają mniej błędów. Jednakże istniejące systemy słabo radzą sobie z analizą video, a w szczególności (mimo że świetnie radzą sobie z rozpoznaniem ludzi na zdjęciach) z powiedzeniem co dana osoba robi na zdjęciu. Co więcej, często do określenia czynności konieczna jest analiza sekwencji obrazów, co nadal stanowi wyzwanie dla tego typu systemów.

Ministerstwo Spraw Wewnętrznych i Administracji wyraziło wstępne zainteresowanie tym tematem w ramach prac przygotowawczych programu INFOSTRATEG.

Cel projektu. Celem projektów realizujących ten temat będzie stworzenie systemu będącego w stanie rozpoznawać co jedna osoba, bądź grupa osób, robi na filmie.

Produkty projektu.

- Zbiór opisanych filmów, zawierających wykonywanie różnych czynności.
- Benchmarki pozwalające określić jakość rozwiązań wymaganych do wdrożenia.
- Przetestowane rozwiązanie.

Korzyści.

- Stworzone systemy mogą zostać wykorzystane do analizy wideo, pochodzących z systemów monitoringu, np. do wykrywania niebezpiecznych sytuacji, przestępstw, czy też rozpoznawania symptomów chorób.
- Rozpoznawanie zachowań ludzkich może być niezbędne przy sterowaniu pojazdami autonomicznymi, operującymi w środowisku, w którym obecni są inni, nieautomatyczni, użytkownicy (piesi, kierowcy, itp.). Rozpoznawanie zachowań może pozwolić na ocenę intencji innych użytkowników dróg.

T11) Weryfikowanie źródeł informacji i detekcja fake newsów

Szacowany budżet: 30 mln PLN.

Sytuacja obecna. Prawdopodobnie największym rozczarowaniem związanym z zapewnieniem prawie wszystkim swobodnego dostępu do informacji, jest szerokie rozpowszechnienie się nieprawdziwych bądź zupełnie mylących informacji. Mamy tutaj na myśli ogólnie pojęte fake newsy, ale także przekonania niezgodne z wiedzą naukową – jak szkodliwość szczepionek, czy wiara w płaską ziemię. Można się spodziewać, że fenomen ten będzie się nasilał i pojawi się więcej błędnych informacji w ogólnie dostępnej sieci. Należy mieć na uwadze, że informacje te często tworzone są po to, by wpłynąć doraźnie na opinię publiczną. Z tego względu konieczne jest stworzenie automatycznych systemów wspomagających wykrywanie tego typu informacji.

Wstępne zainteresowanie tym tematem w ramach prac przygotowawczych programu INFOSTRATEG wyraziło Ministerstwo Spraw Wewnętrznych i Administracji. Jednakże aspekt bezpieczeństwa i weryfikacji informacji zgłaszany był także przez pozostałych interesariuszy.

Cel projektu. Celem projektów realizujących ten temat będzie zbudowanie narzędzi wykrywających fałszywe informacje i pozwalających na oznaczanie takich informacji w wyszukiwarkach. Stworzone systemy mogą być dla przykładu testowane poprzez monitorowanie wiadomości pojawiających się w mediach społecznościowych w okresie dwumiesięcznym. Podstawą oceny będzie porównanie liczby poprawnie znalezionych fałszywych informacji. Systemy powinny jednak być w stanie śledzić wielorakie popularne systemy komunikacji i rozpowszechniania informacji.

Produkty projektu.

- Zbiór danych testowych zawierających przykłady prawdziwych i fałszywych informacji.

- Benchmarki pozwalające określić jakość rozwiązań wymaganych do wdrożenia.
- Przetestowane rozwiązanie.

Korzyści.

- Zadanie to przyczyni się do zmniejszenia negatywnego wpływu fake newsów i ograniczenia ich rozprzestrzeniania.
- Stworzenie takiego systemu może pomóc ograniczyć próby manipulacji opinią publiczną.

T12) Analiza polskich danych medycznych

Szacowany budżet: 50 mln PLN.

Sytuacja obecna. Polska służba zdrowia podlega wręcz rewolucyjnym zmianom dzięki następującej szybkimi krokami cyfryzacji, np. wprowadzenie eRecepty. Aktualnie dane zbierane przez NFZ dotyczą nie tylko wydanych recept, ale także ich realizacji, a także przebiegu i kosztów leczenia. Z czasem przechowywane informacje będą coraz bardziej obszerne. Już teraz widać, że tak szerokie dane stanowią potencjał dla zarządzających systemem zdrowia i pozwalają na dogłębną analizę kosztów, a także na tworzenie systemów przewidywania, czy prób wskazywania grup szczególnie narażonych na niektóre choroby. Jednakże aby tego typu funkcjonalności były dostępne należy przeprowadzić dogłębną analizę tych danych.

Temat ten został zaproponowany w ramach konsultacji z Ministerstwem Zdrowia.

Cel projektu. Celem projektów realizujących ten temat, będzie stworzenie narzędzi pozwalających na statystyczną analizę danych medycznych zbieranych w Polsce. Dane te zawierają między innymi informacje: o wystawionych receptach, zakupionych lekach, zwolnieniach, czy świadczeniach medycznych. Stworzone systemy na podstawie tych połączonych danych powinny być w stanie dostarczać następujące analizy:

- dla danego pacjenta oceniać skuteczność różnych scenariuszy leczenia i alternatywnych leków;
- dla danego pacjenta wyceniać całkowity koszt leczenia;
- dla danego pacjenta przewidywać terminy wykonania badań profilaktycznych;
- dla danej jednostki chorobowej przewidywać skuteczność różnych scenariuszy leczenia i alternatywnych leków;
- przewidywać możliwe przypadki nadużyć.

Produkty projektu.

- Udostępnienie odpowiednio zanonimizowanych danych medycznych zbieranych w Polsce.

- Benchmarki pozwalające określić jakość rozwiązań wymaganych do wdrożenia.
- Przetestowane rozwiązanie.

Korzyści.

- Skuteczne wykrywanie różnych rodzajów nadużyć na szeroką skalę.
- Możliwość skuteczniejszego zarządzania i planowania wydatków przez polską służbę zdrowia.
- Poprawienie skuteczności diagnostyki medycznej poprzez wskazywanie grup pacjentów podwyższonego ryzyka.

5. Sposób interwencji i warunki realizacji projektów w ramach Programu

Program jest realizowany na zasadach określonych w ustawie o NCBR i w towarzyszących aktach wykonawczych. Realizacja Programu polega na finansowaniu projektów i zarządzaniu nimi w sposób zapewniający: osiągnięcie celów oraz zgodność z harmonogramem i planem finansowym.

Konkursy są ogłaszane przez Dyrektora Centrum zgodnie z art. 36 ust. 1 ustawy o NCBR i realizowane zgodnie z obowiązującym w NCBR systemem zarządzania programami strategicznymi. Szczegółowy tryb ogłaszania i rozstrzygania konkursów jest każdorazowo określony w regulaminie konkursu.

Ogólne ramy interwencji, w tym rodzaj zadań objętych finansowaniem oraz intensywność wsparcia, zostały przedstawione w tabeli 5.1., przy czym warunki te będą każdorazowo uszczegółowione w regulaminie konkursu, w sposób adekwatny do zakresu konkursu.

Tabela 5.1. Konstrukcja Programu

Sposób interwencji	Dofinansowanie projektów obejmujących badania naukowe, prace rozwojowe oraz działania obejmujące przygotowanie ich wyników do zastosowania w praktyce.
Wnioskodawcy / Beneficjenci	Jednostki naukowe, przedsiębiorcy, konsorcja składające się z jednostek naukowych, przedsiębiorców oraz innych podmiotów. Katalog Wnioskodawców uszczegóławiany każdorazowo w Regulaminie konkursu
Rodzaje zadań objęte dofinansowaniem	Badania naukowe i prace rozwojowe, przy czym badania podstawowe nie mogą przekraczać 15% całkowitego budżetu projektu. Działania obejmujące przygotowanie wyników badań naukowych i prac rozwojowych do zastosowania w praktyce.
Czas realizacji projektu	do 3 lat

Instrumenty i intensywność wsparcia	Jednostki naukowe - dofinansowanie badań naukowych i prac rozwojowych – do 100% kosztów kwalifikowalnych; - dofinansowanie działań związanych z przygotowaniem wyników badań i prac rozwojowych do zastosowania w działalności gospodarczej – do 100% kosztów kwalifikowalnych; Przedsiębiorcy: pomoc publiczna <u>na badania przemysłowe/aplikacyjne:</u> - małe / mikroprzedsiębiorstwa – do 80% kosztów kwalifikowalnych - średnie przedsiębiorstwa – do 75% kosztów kwalifikowalnych - duże przedsiębiorstwa – do 65% kosztów kwalifikowalnych <u>na prace rozwojowe:</u> - małe/ mikroprzedsiębiorstwa – do 60% kosztów kwalifikowalnych - średnie przedsiębiorstwa - do 50% kosztów kwalifikowalnych - duże przedsiębiorstwa - do 40% kosztów kwalifikowalnych <u>na działania związane z przygotowaniem wyników badań i prac rozwojowych do zastosowania w działalności gospodarczej:</u> - pomoc <i>de minimis</i> – do 90% kosztów kwalifikowalnych, maksymalnie 200 000 euro Inne podmioty, niebędące przedsiębiorcami ani jednostkami naukowymi, posiadające zdolność do wdrożenia rozwiązania będącego wynikiem realizacji Projektu: <u>na badania przemysłowe/aplikacyjne:</u> - do 100% kosztów kwalifikowalnych <u>na prace rozwojowe:</u> - do 100% kosztów kwalifikowalnych <u>na działania związane z przygotowaniem wyników badań i prac rozwojowych do zastosowania w działalności gospodarczej:</u> do 100% kosztów kwalifikowalnych
Katalog kosztów kwalifikowalnych	Określany każdorazowo w Regulaminie konkursu

5.1. Konkursy

Realizacja Programu polega na wyborze projektów w trybie konkursowym, a następnie ich finansowaniu i zarządzaniu nimi w sposób zapewniający osiągnięcie celów Programu. Wykonawcy projektów wybierani są w dwóch typach konkursów:

- na **projekty tematyczne** – obejmujących zakres tematyczny Programu (T1-T12);
- na **projekty zamawiane** – obejmujących zakres tematyczny określony w wyniku skierowanego do podmiotów administracji publicznej naboru na **propozycje tematów badawczych projektów zamawianych**.

W odniesieniu do **propozycji tematów badawczych projektów zamawianych** zgłaszanych przez podmioty administracji publicznej spełnione musi być poniższe wymaganie:

- instytucja zgłaszająca temat powinna określić na jakich danych będą przeprowadzone badania i albo je udostępnić albo opisać w jaki sposób będą one stworzone;

a także wymagania określone już w rozdz. 4:

- jeżeli do prowadzenia przewidzianych badań potrzebne są dane testowe, należy precyzyjnie określić w jaki sposób mogą być zdobyte, bądź wytworzone w ramach realizacji projektów;
- realizacja tematów powinna wymagać użycia nowoczesnych narzędzi uczenia maszynowego, bądź dalszego rozwoju technologii blockchain;
- w tematach ma dominować aspekt tworzenia oprogramowania, choć komponent sprzętowy może w nich występować;
- dla zaproponowanych tematów powinien istnieć jasno określony, ilościowy sposób porównywania między sobą różnych rozwiązań.

Szczegółowe warunki naboru propozycji zostaną określone w Regulaminie naboru propozycji na tematy badawcze projektów zamawianych. Szacunkowy budżet przeznaczony na projekty zamawiane to 360 mln PLN.

Wykonawcy projektów są wybierani w drodze konkursu ogłaszanego przez Dyrektora Centrum zgodnie z art. 36 ust. 1 ustawy o NCBR. Zakłada się, że konkursy na projekty tematyczne będą ogłaszane **raz do roku i będą obejmowały więcej niż jeden temat z zakresu tematycznego Programu**. Konkursy na projekty zamawiane będą ogłaszane nie rzadziej niż raz do roku i będą poprzedzone naborem propozycji na tematy badawcze projektów zamawianych.

Zasadnicze założenia realizacji Programu w celu opracowania optymalnych rozwiązań to:

- realizacja poszczególnych projektów w podziale na etapy (o określonych maksymalnych kosztach realizacji), których efektem ma być osiągnięcie określonego rezultatu/kamienia milowego, co jednocześnie warunkuje finansowanie kolejnego etapu projektu;
- możliwość wyboru kilku projektów, których przedmiotem będzie opracowanie alternatywnych rozwiązań tego samego problemu.

W ramach każdego zakresu tematycznego realizacja projektów będzie podzielona na następujące roczne etapy:

- Faza I Proof of Concept – stworzenie wstępnej wersji systemu, która może zostać poddana ocenie oraz zbieranie danych jeżeli temat tego wymaga;
- Faza II – stworzenie generycznego systemu, np. umiającego obsłużyć wszystkie scenariusze użycia;
- Faza III – ulepszanie powstałego systemu i prace przedwdrozeniowe.

Komitet Sterujący zdefiniuje scenariusze poszczególnych etapów w sposób adekwatny do zakresu tematycznego konkursu. Proponowany mechanizm zakłada konkurowanie proponowanych rozwiązań zarówno na etapie wyboru projektów, jak i ich realizacji. Poniżej przedstawiono założenia liczby projektów dopuszczonych do realizacji kolejnych faz, a także kwoty dofinansowania przypadające na każdą z nich:

- Faza I – minimum 3 projekty po minimum 0,5 mln zł każdy;
- Faza II – minimum 2 projekty po minimum 1 mln zł każdy;
- Faza III – minimum 2 projekty po minimum 2 mln zł każdy.

Kwoty na poszczególne fazy określi Komitet Sterujący w Regulaminie każdego z konkursów w sposób adekwatny do zakresu tematycznego. Oczekiwane rezultaty, warunki wyboru i realizacji projektów zostaną każdorazowo uszczegółowione w Regulaminie konkursu. Regulamin będzie również określał wymagania dotyczące zbieranych danych (tak, aby możliwe było stworzenie na ich podstawie jednego zbioru), a także (na podstawie uzyskanych danych) definiować standardy oceny rozwiązań. W przypadku niespełnienia minimalnej liczby projektów decyzję o dalszej kontynuacji prac nad tematem podejmie Komitet Sterujący na podstawie oceny zaawansowania prac oraz potencjału wykonawców.

Dofinansowanie na realizację projektów będzie przekazywane na podstawie podpisanych umów. Ich rozliczanie będzie się odbywać zgodnie z ustanowioną w NCBR Procedurą *PG1-3: Nadzór nad wykonaniem i finansowaniem projektu w ramach umowy*. Przekazywanie kolejnych transz środków finansowych będzie uzależnione od akceptacji/przyjęcia wyników kolejnych etapów prac, przedstawienia przez wykonawców konkretnych produktów i mierzalnych rezultatów.

6. Sposób monitorowania i oceny realizacji celów Programu

W celu umożliwienia skutecznego monitorowania i oceny stopnia realizacji celów Programu w odniesieniu do celu głównego oraz każdego celu szczegółowego zaproponowano zestaw mierzalnych wskaźników rezultatu, obejmujących kwantyfikowalne wyniki planowane do uzyskania w ramach Programu.

Wskaźniki wpływu (*impact indicators*) mierzą efekty Programu w dłuższej perspektywie czasu i pokazują trwałe zmiany, wykraczające poza bezpośrednie i natychmiastowe rezultaty finansowanych projektów. Tym samym, wskaźniki wpływu można uważać za miernik stopnia realizacji celu głównego Programu. Wskaźniki wpływu będą mierzone w trakcie ewaluacji ex-post wykonanej 3 lata po zakończeniu Programu m.in. na podstawie raportów oraz ankiet dostarczonych przez wykonawców oraz na podstawie zewnętrznej ewaluacji. W tym punkcie analizie zostanie poddana trwałość oraz upowszechnienie rozwiązań wdrożonych w ramach Programu.

Wskaźniki rezultatu (*result indicators*) mierzą bezpośrednie rezultaty projektów, które nastąpiły w wyniku ich realizacji i powinny być możliwe do uchwycenia po zakończeniu realizacji projektów. Wskaźniki rezultatu powinny być logicznie powiązane ze szczegółowymi celami Programu. Wskaźniki rezultatu będą mierzone po zakończeniu projektu, a przed upływem 3 lat od zakończenia Programu, na podstawie informacji dostarczonych przez wykonawców np. raportów końcowych oraz ewaluacji.

Wskaźniki produktu (*output indicators*) są to policzalne, bezpośrednie produkty niezbędne do wytworzenia rezultatów projektów finansowanych w ramach Programu. Wskaźniki produktu będą mierzone w trakcie i po zakończeniu realizacji projektu finansowanego w ramach Programu, na podstawie raportów okresowych oraz raportu końcowego, dostarczonych przez wykonawców.

Docelowe wartości wskaźników pomiaru stopnia osiągnięcia celu głównego i celów szczegółowych Programu wskazano w Tabeli 6.1

Tabela 6.1. Wskaźniki wpływu

Cel główny: Rozwój polskiego potencjału SI poprzez opracowanie rozwiązań wykorzystujących sztuczną inteligencję i blockchain, mających bezpośrednie zastosowanie w praktyce				
Lp	Opis	Wartość bazowa	Wartość docelowa	Rok osiągnięcia
1	Liczba zastosowanych w praktyce rozwiązań opracowanych w ramach Programu	0	15	2029
2	Liczba podmiotów (gospodarczych i publicznych) korzystających z rozwiązań opracowanych w ramach Programu	0	100	2029
3	Liczba osób korzystających z rozwiązań opracowanych w ramach Programu	0	10 000	2029

Tabela 6.2. Wskaźniki rezultatu

Lp	Opis	Wartość bazowa	Wartość docelowa	Rok osiągnięcia
Cel szczegółowy I: Utworzenie zbiorów danych testowych i stworzenie na ich bazie standardów wyboru najlepszych rozwiązań				
1	Liczba zbiorów danych testowych utworzonych w ramach Programu	0	10	2027
Cel szczegółowy II: Rozwijanie polskiego potencjału badawczego w sztucznej inteligencji				
1	Liczba powstałych w ramach Programu publikacji na wiodących konferencjach dot. sztucznej inteligencji*	0	70	2027
2	Wzrost liczby osób ze stopniem doktora zatrudnionych w przedsiębiorstwach uczestniczących w Programie	0	20	2027
Cel szczegółowy III: Znaczące zwiększenie aktywności rynkowej polskich zespołów informatycznych				
1	Liczba utworzonych przez badaczy nowych podmiotów gospodarczych oferujących na rynku konkurencyjne rozwiązania opracowane w ramach Programu	0	12	2027
Cel szczegółowy IV: Zastosowanie sieci neuronowych w robotyce i automatyzacji				
1	Liczba rozwiązań gotowych do produkcji na skalę przemysłową	0	1	2027
Cel szczegółowy V: Stworzenie narzędzi opartych o technologię blockchain przyspieszających rozwój gospodarki cyfrowej				
1	Liczba rozwiązań gotowych do wdrożenia na rynek/ do gospodarki	0	1	2027
Cel szczegółowy VI: Stworzenie rozwiązań opartych o uczenie maszynowe podnoszących jakość produktów/ usług i efektywność procesów				
1	Liczba rozwiązań gotowych do wdrożenia na rynek/ do gospodarki	0	10	2027

* W przypadku uczenia maszynowego oraz technik blockchain jako wiodące konferencje rozumie się następujące wydarzenia:

- Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)
- Association for the Advancement of Artificial Intelligence Conference (AAAI)
- Association for Computational Linguistics Conference (ACL)
- Conference on Computer Vision and Pattern Recognition (CVPR)
- Conference on Empirical Methods in Natural Language Processing (EMNLP)
- Conference on Learning Theory (COLT)
- Conference on Neural Information Processing Systems (NeurIPS)
- Conference on Uncertainty in Artificial Intelligence (UAI)
- Genetic and Evolutionary Computation Conference (GECCO)
- International Conference on Acoustics, Speech, and Signal Processing (ICASSP)
- International Conference on Artificial Intelligence and Statistics (AISTATS)
- International Conference on Autonomous Agents and Multiagent Systems (AAMAS)
- International Conference on Computer Vision (ICCV)
- International Conference on Intelligent Robots and Systems (IROS)
- International Conference on Machine Learning (ICML)
- International Conference on Medical Image Computing & Computer Assisted Intervention (MICCAI)
- International Conference on Robotics and Automation (ICRA)
- International Joint Conferences on Artificial Intelligence (IJCAI)
- Interspeech
- Robotics: Science and Systems (RSS)
- Winter Conference on Applications of Computer Vision (WACV)
- W przypadku bitcoin jako wiodące konferencje rozumie się następujące wydarzenia:
- IEEE Symposium on Security and Privacy (S&P)
- ACM SIGSAC Conference on Computer and Communications Security, CCS
- USENIX Security Symposium
- International Cryptology Conference - CRYPTO
- Advances in Cryptology - EUROCRYPT
- Advances in Cryptology - ASIACRYPT
- Network and Distributed System Security Symposium (NDSS)
- Financial Cryptography and Data Security
- ACM SIGACT-SIGOPS Symposium on Principles of Distributed Computing (PODC)

Tabela 6.3. Wskaźniki produktu

Lp	Opis	Wartość bazowa	Wartość docelowa	Rok osiągnięcia
1	Liczba jednostek administracji państwowej/ firm, które wezmą udział w tworzeniu danych testowych	0	10	2025
2	Liczba doktoratów uzyskanych w związku z udziałem w realizacji projektów w ramach Programu	0	15	2025
3	Liczba współautorskich publikacji jednostek naukowych i przedsiębiorców, dotyczących wyników prac B+R uzyskanych w ramach Programu (w czasopismach i konferencjach objętych listą ministerialną o wartości punktowej co najmniej 100)	0	15	2025
4	Liczba jednostek administracji państwowej/ firm, które określą minimalną jakość benchmarków/ środowisk testowych	0	12	2025
5	Liczba przetestowanych pilotażowo maszyn/ robotów opartych na uczeniu maszynowym	0	1	2025

6	Liczba opracowanych rozwiązań opartych na uczeniu maszynowym	0	11	2025
7	Liczba opracowanych rozwiązań opartych na technologii blockchain	0	1	2025

6.1. Ewaluacja

W trakcie realizacji Programu będzie prowadzona jego ewaluacja, w szczególności w celu rozstrzygnięcia, czy kontynuacja Programu prowadzi do osiągnięcia jego celów oraz czy jest on zgodny z celami polityki naukowej państwa i polityki wspierania innowacyjności. Po zakończeniu realizacji Programu, przeprowadzona będzie ewaluacja mająca na celu w szczególności ocenę stopnia osiągnięcia jego celów, a w przypadku ich nieosiągnięcia określenie przyczyn niepowodzenia. Proces ewaluacji będzie realizowany zgodnie z obowiązującą w NCBR Procedurą PG2-2: *Ewaluacja programu*.

7. Określenie ryzyka dla osiągnięcia celów Programu

Ryzyka związane z nieosiągnięciem celów Programu mają charakter zewnętrzny (niezależny od NCBR) oraz wewnętrzny – w tym przypadku możliwe jest podjęcie działań minimalizujących ryzyka wewnętrzne.

Zewnętrzne:

- **zmiany okoliczności zewnętrznych** - na skutek trudnych do przewidzenia czynników politycznych, ekonomicznych czy społecznych - nowe okoliczności mogą uniemożliwić realizację zamierzonych działań. Przykładem takiego czynnika może być polityczny odgórny zakaz rozwoju pewnego rodzaju rozwiązań SI;
- **brak realnego odbiorcy rozwiązań** – brak ścisłej współpracy pomiędzy wykonawcą i instytucją, mającą potencjał do wdrożenia danego rozwiązania, może doprowadzić do sformułowania propozycji nie mających realnego odbiorcy;
- **brak odpowiedniego zbioru danych** – niektóre zastosowania SI mogą wymagać ogromnej liczby dobrze opisanych danych do skutecznego i efektywnego wytrenowania algorytmów. Istnieje ryzyko braku możliwości przygotowania tak dużego zbioru danych w zadanym terminie;
- **ryzyka prawne związane z własnościami danych** – istnieje ryzyko, że wykorzystane dane mogą wymagać pozyskania wielu skomplikowanych zgód (ze względu na RODO), w tym zgód konsumentów, co może znacznie utrudnić realizację zadania;
- **niski potencjał wdrożeniowy rozwiązań** – w przypadku zbyt teoretycznego podejścia do problemu, wynikającego z nieuwzględniania złożonych realiów, może dojść do formułowania rozwiązań niekompatybilnych z dostępnymi systemami, ekonomicznie nieopłacalnych, nie mających potencjału wdrożeniowego w obecnych realiach;

Wewnętrzne:

- **rozmyta odpowiedzialność** – bez aktywnej roli Komitetu Sterującego oraz koordynatora Programu może dojść do rozmycia odpowiedzialności;

działania mitygujące: określenie zadań i kompetencji Komitetu Sterującego oraz koordynatora w dokumentach wdrożeniowych: Regulaminie pracy Komitetu Sterującego, procedurach wewnętrznych NCBR oraz Regulaminach konkursów.

- **brak skutecznej promocji Programu** – przy braku skutecznej promocji Programu istnieje ryzyko małego zainteresowania Programem;
działania mitygujące: opracowanie i wdrożenie przez NCBR planu promocji Programu.

- **niedopasowany czas trwania projektu (harmonogram) do czasu wdrożenia lub zły moment ogłoszenia Konkursu** – ryzyko niedopasowania czasu trwania realizacji projektu w stosunku do przyjętego czasu na osiągnięcie zakładanych celów Programu (np. ogłoszenie konkursu dla rolnictwa w nieodpowiednim okresie roku może skutkować utrudnionym okresem dla zebrania danych do automatyzacji rolnictwa);
działania mitygujące: uwzględnienie przez Komitet Sterujący krytycznych czynników w szczegółowym harmonogramie realizacji Programu.

- **dezaktualizacja agendy badawczej** – w przypadku zbyt długiego okresu czasu pomiędzy zdiagnozowaniem problemu a wypracowaniem rozwiązań może dojść do dezaktualizacji agendy badawczej;
działania mitygujące: aktualizacja zakresu tematycznego Programu, uwzględnienie trybu projektów zamawianych, stanowiących odpowiedź na problemy niezidentyfikowane w zakresie tematycznym Programu.

- **źle dokonana wymiana danych** – brak transferowalności i kompatybilności danych pomiędzy różnymi systemami i rozwiązaniami SI, wynikający z niedostatecznie dokładnie wyspecyfikowanego standardu danych;
działania mitygujące: określenie przez Komitet Sterujący zasad wymiany danych pomiędzy zespołami tak, aby zachować ich transferowalność i kompatybilność.

- **utrudnione porównanie skuteczności rozwiązania** – szczególnie w przypadku porównywania wyników pomiędzy różnymi zastosowaniami (np. porównanie wdrożenia asystenta lekarza pediatry wobec lekarza ortopedy);
działania mitygujące: określenie przez Komitet Sterujący zasad porównywania rozwiązań stworzonych przez różne zespoły.

- **źle oszacowane koszty Konkursu** – z uwagi na innowacyjny charakter Programu możliwe jest zróżnicowanie w zakresie wysokości kosztów pomiędzy obszarami. Istnieje ryzyko przeszacowania lub niedoszacowania kosztów realizacji projektów;
działania mitygujące: elastyczne podejście Komitetu Sterującego oraz NCBR, umożliwiające dostosowanie do zaistniałej sytuacji z możliwością relokacji szacunkowych kosztów pomiędzy tematami w zakresie tematycznym Programu.

- **brak możliwości zapewnienia odpowiedniej prywatności danych** – część badań będzie wymagała dostępu do danych medycznych i istnieje ryzyko, że w przypadku, np. wywiadów medycznych odpowiednia anonimizacja będzie trudna do zrealizowania.
działania mitygujące: określenie przez Komitet Sterujący we współpracy z NCBR odpowiednich zasad anonimizacji danych i zarządzania nimi.

Zarządzanie ryzykiem w Programie prowadzone będzie przez koordynatora Programu według obowiązującej w NCBR *Procedury nr PZ3-1: Zarządzanie Ryzykiem*.

8. Harmonogram realizacji Programu

Program ustanawiany jest na okres 2020 - 2029 z możliwością wydłużenia lub skrócenia czasu trwania Programu. Harmonogram realizacji programu obejmuje trzy etapy:

1. przygotowania do uruchomienia Programu;
2. wdrażanie Programu;
3. ewaluację ex-post Programu.

Etap przygotowań do uruchomienia Programu trwa około 3 miesięcy od zatwierdzenia Programu przez Ministra Nauki i Szkolnictwa Wyższego i obejmuje:

- a) powołanie koordynatora Programu;
- b) powołanie Komitetu Sterującego;
- c) ustalenie szczegółowego zakresu tematycznego pierwszego konkursu i określenie alokacji środków przez Komitet Sterujący;
- d) przygotowanie dokumentacji I konkursu;
- e) zaopiniowanie przez Radę NCBR Regulaminu I konkursu.

Ramowy harmonogram realizacji Programu zakłada:

1. **2020** – etap przygotowań do uruchomienia Programu;
2. **2020–2027**: wdrażanie Programu
 - **2020–2022**: organizację konkursów, wybór projektów;
 - **2022–2025**: realizację wybranych projektów, analizę i ewaluację śródkresową Programu;
 - **2025 – 2027**: analizę wyników zakończonych projektów.
3. **2029** – ewaluację ex-post Programu

Szczegółowy harmonogram realizacji Programu proponuje Komitet Sterujący z uwzględnieniem budżetu NCBR na rok bieżący oraz kolejne lata realizacji Programu.

9. Budżet i plan finansowy Programu oraz źródła finansowania

Budżet Programu wynosi **840 mln PLN** (w tym 360 mln PLN na projekty zamawiane) i pochodzi z dotacji celowej na realizację strategicznych programów badań naukowych i prac rozwojowych, o której mowa w art. 46 ust. 1 pkt 1 ustawy o NCBR oraz ze środków pozabudżetowych, włączając środki przedsiębiorców oraz środki innych instytucji działających w obszarach Programu. W związku z możliwością wydłużenia / skrócenia czasu trwania Programu przewiduje się możliwość zmiany wielkości budżetu w trakcie realizacji Programu.

Koszty zarządzania Programem, w tym wynagrodzenia pracowników NCBR zaangażowanych we wdrażanie Programu, koszty oceny wniosków o dofinansowanie wykonywanych przez niezależnych ekspertów, koszty związane z działalnością Komitetu Sterującego, będą pochodziły z dotacji podmiotowej na pokrycie bieżących kosztów zarządzania realizowanymi przez NCBR zadaniami, o której mowa w art. 46 ust. 1 pkt 2 ustawy o NCBR. Zakłada się, że koszty zarządzania Programem nie przekroczą 5% budżetu NCBR przeznaczonego na finansowanie projektów w ramach Programu.

Co najwyżej 95% środków przeznaczonych na dofinansowanie projektów w ramach Programu będzie pochodziło z budżetu NCBR. Pozostałą część dofinansowania (co najmniej 5%) muszą zapewnić środki pochodzące od przedsiębiorców lub instytucji działających w obszarach Programu.

Proponowany podział dofinansowania projektów w ramach Programu będzie ustalany wspólnie przez Dyrektora Centrum i Komitet Sterujący Programu.

10. System realizacji i zarządzanie Programem

Nadzór nad realizacją Programu sprawuje Dyrektor Centrum. Strukturę zarządzania realizacją Programu określa *Procedura PG1-1C Uruchamianie i realizacja programu strategicznego badań naukowych i prac rozwojowych (nie dotyczy programu na rzecz obronności i bezpieczeństwa państwa)* i obejmuje w szczególności Dyrektora Centrum, Radę Centrum, Komitet Sterujący i koordynatora.

Sposób powołania i zadania Komitetu Sterującego oraz koordynatora Programu są określone w *Procedurze PG1-1C Uruchamianie i realizacja programu strategicznego badań naukowych i prac rozwojowych (nie dotyczy programu na rzecz obronności i bezpieczeństwa państwa)*.

Realizacja Programu obejmuje m.in. ogłaszanie konkursów, wybór beneficjentów, finansowanie projektów obejmujących badania naukowe, prace rozwojowe oraz działania związane z przygotowaniem wyników badań i prac rozwojowych do zastosowania w praktyce. Wykonawcy projektów są wybierani w drodze konkursu ogłaszanego przez Dyrektora Centrum zgodnie z art. 36 ust. 1 ustawy o NCBR. System zarządzania procesami w ramach Programu jest zgodny z procedurami NCBR - najważniejsze z nich wskazano w Tabeli 10.1.

Mając na względzie specyfikę zakresu tematycznego Programu, rozwiązanie postawionych tematów wiąże się z kilkoma scenariuszami opracowania rezultatów w zależności od dostępności danych:

1. opracowanie rozwiązania nie wymaga dostępu do zbiorów danych;
2. do opracowania i przetestowania rozwiązania konieczne są dane, a ich pozyskanie stanowi część projektu badawczego;
3. dane niezbędne do opracowania rozwiązania są w dyspozycji konkretnego podmiotu (publicznego), na którego potrzeby odpowiada opracowywane rozwiązanie.

W przypadku 2. i 3. Regulamin konkursu dla danego tematu powinien zawierać opis standardu danych, który musi być wspólny dla realizujących tematy. Natomiast, w 3im przypadku Komitet Sterujący powinien zagwarantować, że odpowiednie dane zostaną przygotowane i udostępnione na cel konkursu.

Zakłada się, że udostępniane dane będą zawsze podlegały anonimizacji, tzn. zostaną z nich usunięte dane osobowe, a dla przykładu głos - zostanie elektronicznie zmieniony. Wstępne analizy istniejących procedur wskazują, że udostępnienie zanonimizowanych danych powinno być możliwe, np. w ramach Forum eZdrowia w 2019 roku NFZ udostępnił uczestnikom hackatonu zanonimizowane dane medyczne znaczącej części społeczeństwa. Ze względu na różnorodność danych, których dotyczy program, sposób udostępniania danych będzie każdorazowo określony w regulaminie konkursu. Co więcej, jeżeli nie będzie przeciwwskazań, to dane zostaną

udostępnione przy użyciu ogólnie dostępnych repozytoriów danych, np. Otwarte Dane, czy EU Open Data Portal.

Na koniec trzeba zaznaczyć, że realizowane projekty będą przechodziły analizę etyczną zgodną z ogólnymi zasadami NCBR.

Tabela 10.1: Procedury związane z realizacją i zarządzaniem programami strategicznymi

Procedura	Proces
PG1-1C	Uruchamianie i realizacja programu strategicznego badań naukowych i prac rozwojowych (nie dotyczy programu na rzecz obronności i bezpieczeństwa państwa)
PG1-1D	Wprowadzanie zmian do strategicznego programu badań naukowych i prac rozwojowych (nie dotyczy programu na rzecz obronności i bezpieczeństwa państwa)
PG1-2	Wybór wykonawców projektów
PG1-3	Nadzór nad wykonaniem i finansowanie projektu w ramach umowy
PG1-4	Monitorowanie programu/przedsięwzięcia
PG2-1	Dostarczanie ocen
PG2-2	Ewaluacja programu
PG2-3	Postępowanie w sprawie rozpatrywania odwołań od decyzji Dyrektora NCBR w sprawie przyznania środków finansowych na wykonanie projektów, odwołań od promesy finansowania, odwołań od innych decyzji Dyrektora NCBR oraz zażaleń od postanowień wydawanych przez Dyrektora NCBR
PG2-4	Postępowanie w sprawie rozpatrywania zarzutów podnoszonych w zażaleniach od postanowień o odmowie wszczęcia postępowania po ocenie wstępnej (formalnej) Dyrektora Centrum kończących postępowanie w sprawie rozpatrzenia wniosku
PG2-5	Planowanie i realizacja kontroli projektów przez NCBR
PG2-6	Windykacja należności

11. Matryca logiczna Programu

WYZWANIA DO ZAADRESOWANIA		ZAŁOŻENIA PROGRAMU STRATEGICZNEGO INFOSTRATEG						
Problem/ szansa	Odpowiedź na potrzebę/ szansę	Cel główny	Cel szczegółowy	Korzyści	Obszary	Wskaźniki produktu	Wskaźniki rezultatu	Wskaźniki
Problem 1: niedostateczny poziom organizacji (np. w zakresie kategoryzacji, opisania czy oczyszczenia), integracji i analizy danych, uniemożliwiający ich pełne i optymalne wykorzystanie w procesie wytrenowania sztucznej inteligencji, brak danych testowych .	Otwarcie i udostępnienie zbiorów danych, które nazwane są paliwem dla SI.	Celem głównym programu jest rozwój polskiego potencjału SI poprzez opracowanie rozwiązań wykorzystujących sztuczną inteligencję i blockchain, mających bezpośrednie zastosowanie w praktyce.	Cel szczegółowy I: Utworzenie zbiorów danych testowych i stworzenie na ich bazie standardów wyboru najlepszych rozwiązań.	Utworzenie danych testowych, utworzenie standardów / uniwersalnych narzędzi- ułatwi i zmniejszy koszt tworzenia nowych rozwiązań dla powiązanych problemów.	T1, T2, T3, T4, T6, T7, T8, T10, T11 oraz T12	1. Liczba jednostek administracji państwowej/ firm, które wezmą udział w tworzeniu danych testowych.	I.1. Liczba zbiorów danych testowych utworzonych w ramach Programu.	1. Liczba zastosowanych w praktyce rozwiązań opracowanych w ramach Programu 2. Liczba podmiotów (gospodarczych i publicznych) korzystających z rozwiązań opracowanych w ramach Programu 3. Liczba osób korzystających z rozwiązań opracowanych w ramach Programu
Problem 2: dość istotny polski potencjał badawczy w dziedzinie SI nie przekłada się na działalność B&R, długi cykl badań nieprzystający do oczekiwań rynku.	Postawienie na rozwój kompetencji i aktywację potencjału badawczego SI, np. poprzez wspieranie realizacji doktoratów wdrożeniowych w dziedzinie SI, maksymalne skracanie cyklu prac badawczych od idei do <i>proof of concept</i> ; pierwszeństwo w wykorzystaniu rozwiązań chmurowych.		Cel szczegółowy II: Rozwijanie polskiego potencjału badawczego w sztucznej inteligencji.	Więcej innowacyjnych rozwiązań SI w krótszym czasie dzięki wzrostowi intensywności współpracy nauki z biznesem; wzrost dostępności wiedzy eksperckiej (doktorantów-praktyków) w zakresie SI.	Wszystkie	2. Liczba doktoratów uzyskanych w związku z udziałem w realizacji projektów w ramach Programu. 3. Liczba współautorskich publikacji jednostek naukowych i przedsiębiorców, dotyczących wyników prac B+R uzyskanych w ramach Programu (w czasopiśmie i konferencjach objętych listą ministerialną o wartości punktowej co najmniej 100).	II.1. Liczba powstałych w ramach Programu publikacji na wiodących konferencjach dot. sztucznej inteligencji. II.2. Wzrost liczby osób ze stopniem doktora zatrudnionych w przedsiębiorstwach uczestniczących w Programie.	
Problem 3: Brak generycznych rozwiązań, standardów, narzędzi, miar odniesienia (benchmarks), które mogłyby przyspieszyć rozpowszechnienie	Stworzenie podstawowej bazy, oprogramowania, know-how i metodologii, powiązanych z uczeniem maszynowym jako bazy do budowy dalszych/ rozpowszechnienie		Cel szczegółowy III: Znaczące zwiększenie aktywności rynkowej polskich	Zwiększenie aktywności rynkowej polskich zespołów informatycznych oraz firm.	Wszystkie	4. Liczba jednostek administracji państwowej/ firm, które określą minimalną jakość benchmarków/ środowisk testowych.	III.1. Liczba utworzonych przez badaczy nowych podmiotów gospodarczych oferujących na rynku	

stosowania sztucznej inteligencji.	bardziej zaawansowanych produktów.		zespołów informatycznych.				rozwiązania opracowane w ramach Programu.	
Problem/ szansa	Odpowiedź na potrzebę/ szansę		Cel częściowy	Korzyści gospodarcze	Obszary tematyczne	Wskaźniki produktu	Wskaźniki rezultatu	
Szansa 1: wykorzystanie SI do wzrostu wykorzystania i efektywności robotów.	Rozwój robotyki opartej o sieci neuronowe.		Cel szczegółowy IV: zastosowanie sieci neuronowych w robotyce i automatyzacji	Wzrost efektywności i jakości produkcji, obniżenie kosztów wytwarzania.	T5	5. Liczba przetestowanych pilotażowo maszyn/ robotów opartych na uczeniu maszynowym	IV.1. Liczba rozwiązań gotowych do produkcji na skalę przemysłową.	
Szansa 2: wykorzystanie technologii blockchain jako akceleratora rozwoju gospodarki cyfrowej.	Opracowanie narzędzi, standardów opartych o technologię blockchain.		Cel szczegółowy V: stworzenie narzędzi opartych o technologię blockchain przyspieszających rozwój gospodarki cyfrowej	Stosowanie eKontraktów oznaczałoby dla firm oszczędności na kosztach prawnych oraz pozwoliłoby zminimalizować potencjalne spory.	T9	6. Liczba opracowanych rozwiązań opartych na technologii blockchain.	V.1. Liczba rozwiązań gotowych do wdrożenia na rynek/ do gospodarki.	
Szansa 3: rosnący potencjał sztucznej inteligencji- są obszary, gdzie komputer szybciej i lepiej wykona zadanie/ rozwiąże problem.	Opracowanie praktycznych rozwiązań opartych o uczenie maszynowe, gdzie czasochłonne, zasobochłonne zadania mogą być szybciej i bezbłędnie realizowane przez maszyny.		Cel szczegółowy VI: stworzenie rozwiązań opartych o uczenie maszynowe podnoszących jakość produktów/ usług i efektywność procesów	Obniżenie kosztów procesów, wzrost jakości usług/ produktów, eliminacja błędów, poprawa skuteczności procesów, skrócenie czasu podejmowania decyzji.	T1, T2, T3, T4, T6, T7, T8, T9, T10, T11 oraz T12	7. Liczba opracowanych rozwiązań opartych na uczeniu maszynowym.	VI.1. Liczba rozwiązań gotowych do wdrożenia na rynek/ do gospodarki.	