QA-OnkoBot S1: Wprowadzenie do Koncepcji Ocenienia Jakości Odpowiedzi & Dialogu Specyfikacji

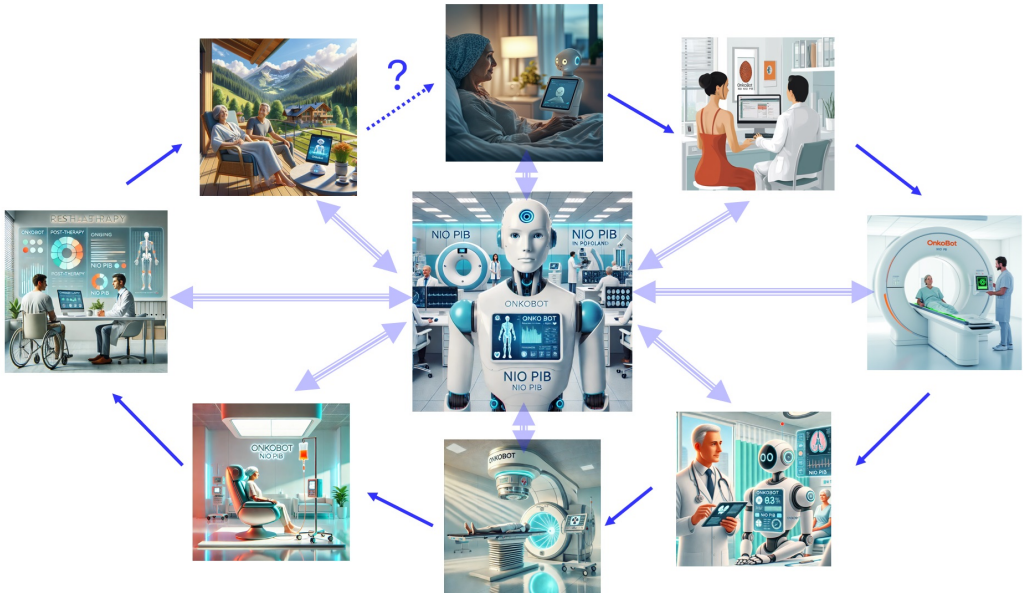
dr hab. A. Jankowski, prof. UWM

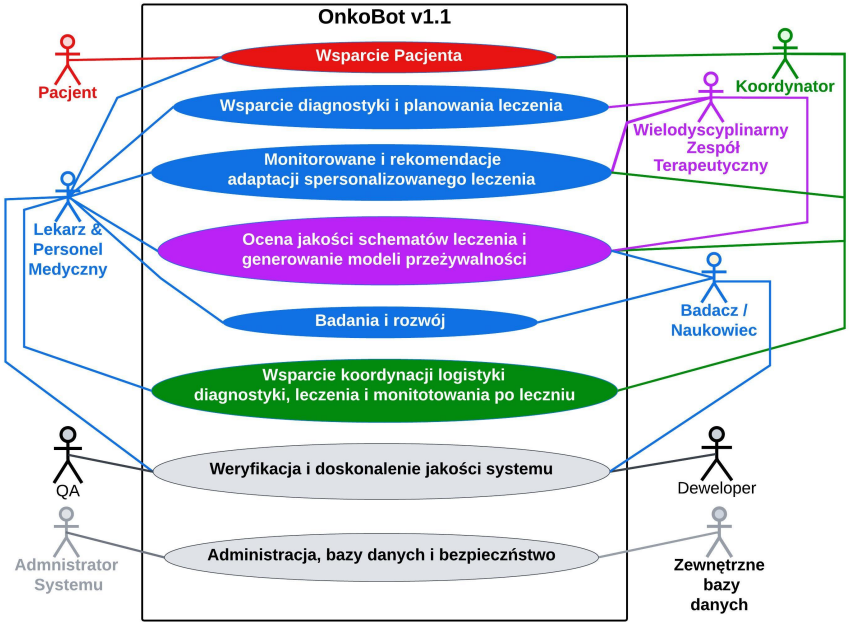
Luty 2025

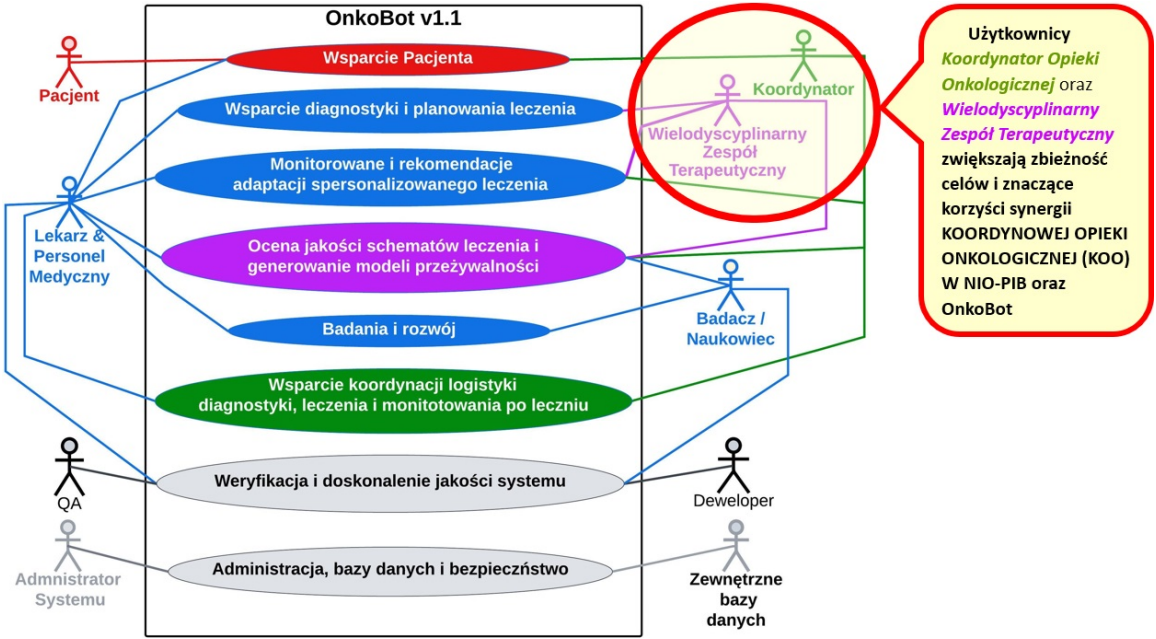
1. Wizja funkcjonalności OnkoBot
2. Kryteria Ocen
   - Zgodność z wiedzą i aktualność
   - Wyczerpująca informacja
   - Personalizacja i jasność komunikacji
   - Empatia i klimat rozmowy
   - Pogłębianie dialogu
   - Przejrzystość źródeł wiedzy
3. Skala Ocen
   - Skala Ocen Pozytywnych
   - Ocena Neutralna
   - Skala Ocen Negatywnych
4. Podstawowe Statystyki Ocen
5. Formularz Zbierania Ocen od Użytkowników
6. Podsumowanie Kryteriów Ocen & Podstawowych Statystyk Oceny Użytkowej Jakości Pytań
7. Wstępna Architektura Pakietów Użytkowych QA-OnkoBot

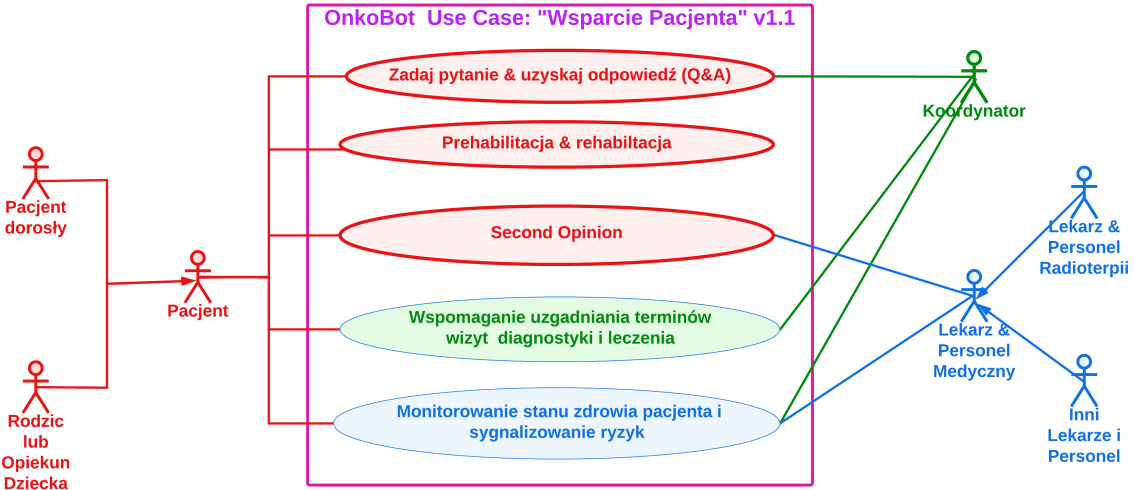
Kontekst & Wprowadzenie

- Budujemy zaawansowany system (LLM/RAG, grafowy oraz reprezentacji wnioskowań medycznych na bazie ontologii) do zastosowań w onkologii o nazwie OnkoBot.
- Jednym z kluczowych czynników jest zapewnienie jakości OnkoBot. Realizację tego celu mają realizować specjalne podprzypadki użycia w ramach modułu QA-OnkoBot.
- W ramach projektu planujemy stosować zarówno ręczne testowanie przez ekspertów oraz potencjalnych użytkowników, jak i również umożliwić automatyczne wsparcie generowanie i oceneniania pytań, odpoiwedzi i dialogów
- Scenariusze testów będą dążyły do identyfikacji i systematycznego poprawiania jakości OnkoBot, w taki sposób aby prawdopodobieństwo usterek systemu OnkoBot minimalizować do zera.
- Niniejsza prezentacja przedstawia wstępną propozycję specyfikacji funkcjonalności QA-OnkoBot.
- Wraz z rozwojem systemu OnkoBot, będzie również doskonalony system QA-OnkoBot.

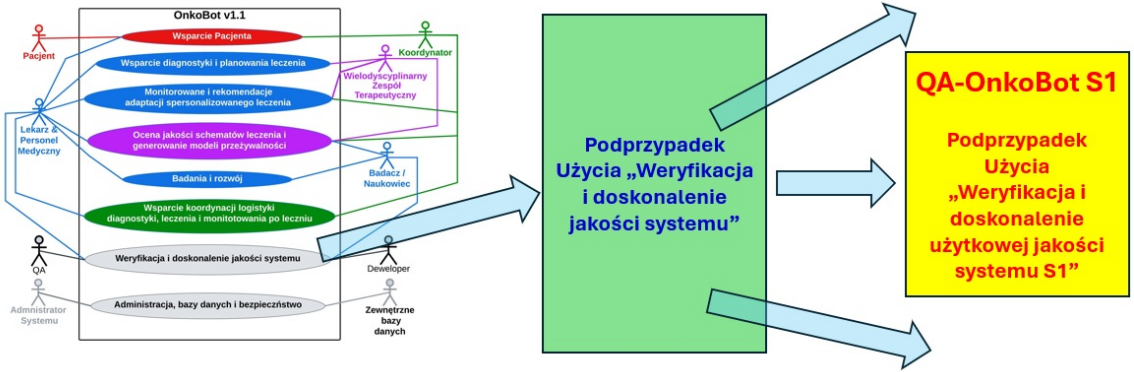

?
NIO PIB
NIO PIB
IN POTSDAM
ONKOBOT
ONKO BOT
NIO PIB
NIO PIB
OnkoBot
ONKOBOT
NIO PIB
ONKOBOT
ONKOBOT
NIO PIB
ONKOBOT
83%
NIO PIB


OnkoBot v1.1
Pacjent
Wsparcie Pacjenta
Koordynator
Wsparcie diagnostyki i planowania leczenia
Wielodyscyplinarny Zespół Terapeutyczny
Monitorowane i rekomendacje adaptacji spersonalizowanego leczenia
Lekarz & Personel Medyczny
Ocena jakości schematów leczenia i generowanie modeli przeżywalności
Badania i rozwój
Badacz / Naukowiec
Wsparcie koordynacji logistyki diagnostyki, leczenia i monitotowania po leczniu
QA
Weryfikacja i doskonalenie jakości systemu
Deweloper
Admnistrator Systemu
Administracja, bazy danych i bezpieczństwo
Zewnętrzne bazy danych


OnkoBot v1.1
Pacjent
Wsparcie Pacjenta
Wsparcie diagnostyki i planowania leczenia
Monitorowane i rekomendacje adaptacji spersonalizowanego leczenia
Lekarz & Personel Medyczny
Ocena jakości schematów leczenia i generowanie modeli przeżywalności
Badania i rozwój
Badacz / Naukowiec
Wsparcie koordynacji logistyki diagnostyki, leczenia i monitotowania po leczniu
QA
Weryfikacja i doskonalenie jakości systemu
Deweloper
Admnistrator Systemu
Administracja, bazy danych i bezpieczństwo
Zewnętrzne bazy danych
Koordynator
Wielodyscyplinarny Zespół Terapeutyczny
Użytkownicy Koordynator Opieki Onkologicznej oraz Wielodyscyplinarny Zespół Terapeutyczny zwiększają zbieżność celów i znaczące korzyści synergii KOORDYNOWEJ OPIEKI ONKOLOGICZNEJ (KOO) W NIO-PIB oraz OnkoBot


OnkoBot Use Case: "Wsparcie Pacjenta" v1.1
Zadaj pytanie & uzyskaj odpowiedź (Q&A)
Prehabilitacja & rehabiltacja
Second Opinion
Wspomaganie uzgadniania terminów wizyt diagnostyki i leczenia
Monitorowanie stanu zdrowia pacjenta i sygnalizowanie ryzyk
Pacjent dorosły
Pacjent
Rodzic lub Opiekun Dziecka
Koordynator
Lekarz & Personel Radioterpii
Lekarz & Personel Medyczny
Inni Lekarze i Personel

Co to jest QA-OnkoBot S1



Faza / Faza.Etap	Makieta	Prototyp 1	Prototyp 2 & BM	Pilot & Badania Medyczne	Wdrożenie & Rozwój	Utrzymanie & Rozwój
F1.E1	S1, S2	S1, S2	S1	S1		
F1.E2	S3	S3	S2	S1-S2		
F2.E1	S4	S4	S3	S1-S3		
F2.E2	S5	S5	S4	S1-S4		
F3.E1	S0, S6	S6	S5	S1, S2, S3, S4,S5		
F3.E2	S7	S0, S7	S6	S3, S4, S5, S6	S1, S2	
F3.E3	S8	S0, S8	S7	S4, S5, S6, S7	S3	S1, S2
F3.E4	S9	S9	S0, S8	S5, S5, S6, S7,S8	S4	S3
F3.E5	S10	S10	S9	S6, S6, S7, S8, S9	S5	S4
F4	S11	S11	S10	S7, S8, S9, S10	S6	S5
F5	S12	S12	S11	S8, S9, S10, S11	S7	S6
F6	S13	S13	S12	S9, S10, S11, S12	S8	S7
F7			S13	S10, S11, S12,S13	S9	S8
F8				S11, S12, S13	S10	S9
F9.E1				S12, S13	S11	S10
F9.E2				S13	S12	S11
F10.E1					S13	S12
F10.E2						S13

Wsparcie Pacjenta Radioterapii	**S0** = Podsystem integracji rozbudowywanego systemu obejmujący interfejsy między lokalnymi bazami danych oraz centralne bazy danych. **S1** = Sys0_0 + Obsługa przez dedykowany model LLM/RAG pytań i odpowiedzi rodziców/opiekunów dzieci poddawanych radioterapii w zakresie najczęściej zadawanych pytań, ze szczególnym uwzględnieniem wsparcia w zakresie "second opinion". **S2** = S1 + obsługa przez dedykowany model LLM/RAG pytań i odpowiedzi dorosłych pacjentów radioterapii w zakresie najczęściej zadawanych pytań, ze szczególnym uwzględnieniem wsparcia w zakresie "second opinion". **S3** = S2 + obsługa pytań i odpowiedzi dotyczących prehabilitacji onkologicznej i rehabilitacji onkologicznej dla rodziców dzieci poddawanych radioterapii. **S4** = S3 + obsługa pytań i odpowiedzi dotyczących prehabilitacji onkologicznej i rehabilitacji onkologicznej dla dorosłych pacjentów radioterapii.
Wsparcie Lekarza Radioterapii	**S5** = S4 + wsparcie radioterapeutów w zakresie wiedzy na temat aktualnych wytycznych dotyczących diagnostyki i leczenia radioterapeutycznego oraz ich kojarzenia z leczeniem systemowym, zarówno na poziomie lokalnym (krajowym), jak i międzynarodowym. **S6** = S5 + wsparcie radioterapeutów w zakresie analizy obrazów diagnostyki obrazowej **S7** = S6 + wsparcie radioterapeutów w zakresie planowania i optymalizacji spersonalizowanego leczenia promieniowaniem **S8** = S7 + integracja wiedzy klinicznej w placówce (lub sieci placówek) w celu stworzenia zintegrowanej bazy wiedzy o pacjentach, skuteczności leczenia i wsparcia monitorowania stanu zdrowia po radioterapii.
Wsparcie Pacjenta NIO PIB	**S9** = S8 + Obsługa pytań pacjentów NIO PIB zakresie „second opinion” oraz prehabilitacji i rehabilitacji onkologicznej
Wsparcie Lekarza NIO PIB	**S10** = S9 + Wsparcie onkologów w zakresie wiedzy na temat aktualnych wytycznych dotyczących diagnostyki i leczenia onkologicznego oraz ich kojarzenia z leczeniem systemowym, zarówno na poziomie lokalnym (krajowym), jak i międzynarodowym **S11** = S10 + Wsparcie onkologów NIO PIB w zakresie analizy obrazów diagnostyki obrazowej **S12** = S11 + Wsparcie onkologów w zakresie planowania i optymalizacji spersonalizowanego leczenia onkologicznego w NIO PIB **S13** = S12 + Integracja wiedzy klinicznej w NIO PIB (lub sieci placówek) w celu stworzenia zintegrowanej bazy wiedzy o pacjentach, skuteczności leczenia i wsparcia monitorowania stanu zdrowia po terapii onkologicznej


Uruchomienie projektu i prototypy „Wsparcia Pacjenta” w Zakładzie Radioterapii (F1)
„Wsparcie Pacjenta” w Zakładzie Radioterapii (F2-F3)
Wsparcie lekarzy w Zakładzie Radioterapii (F2-F4)
„Wsparcie Pacjenta” w całym NIO PIB (F3-F6)
Wsparcie lekarzy w całym NIO PIB (F3-10)

Zgodność z wiedzą i aktualność: Definicja

Definicja

Treść odpowiedzi musi być zgodna z aktualną wiedzą medyczną oraz praktyką kliniczną. Agent powinien:

- Unikać podawania informacji nieprawdziwych, niepełnych lub przestarzałych
- Bazować na najnowszych wytycznych, badaniach naukowych i literaturze medycznej
- Wskazywać datę ostatniej aktualizacji wykorzystywanych wytycznych

Zgodność z wiedzą i aktualność: Przykłady

Przykłady oceny

- Czy odpowiedź uwzględnia najnowsze zalecenia w leczeniu zgodnie z wytycznymi ESMO lub NCCN?
- Czy agent zaznacza, kiedy ostatnio aktualizowano cytowane wytyczne?
- Czy agent wskazuje na różnice między szkołami medycznymi w podejściu do leczenia?
- Czy informacje o skuteczności terapii są poparte aktualnymi badaniami klinicznymi?

Wyczerpująca informacja: Definicja

Definicja

Odpowiedź powinna zawierać wszystkie kluczowe informacje potrzebne do zrozumienia problemu i podjęcia decyzji:

- Unikanie nadmiernej szczegółowości utrudniającej zrozumienie
- Prezentacja informacji od najważniejszych do szczegółowych
- Zachowanie logicznej struktury przekazu

Wyczerpująca informacja: Przykłady

Przykłady oceny

- Czy agent wyjaśnił wszystkie dostępne opcje leczenia, uwzględniając ich dostępność?
- Czy najważniejsze informacje zostały przedstawione na początku?
- Czy agent uwzględnił potencjalne skutki uboczne i przeciwwskazania?
- Czy informacja zawiera aspekty praktyczne?

Personalizacja i jasność komunikacji: Definicja

Definicja

Agent powinien dostosowywać przekaz do odbiorcy:

- Dostosowanie języka do poziomu wiedzy użytkownika
- Uwzględnienie indywidualnych danych pacjenta
- Weryfikacja zrozumienia przekazanych informacji
- Odpowiednie wyjaśnianie terminologii medycznej

Personalizacja i jasność komunikacji: Przykłady

Przykłady oceny

- Czy agent pyta o poziom zaawansowania medycznej wiedzy?
- Czy terminy medyczne są odpowiednio wyjaśniane?
- Czy agent weryfikuje zrozumienie przekazanych informacji?
- Czy odpowiedź uwzględnia specyficzne uwarunkowania pacjenta?

Empatia i klimat rozmowy: Definicja

Definicja

Agent powinien odpowiadać w sposób empatyczny:

- Uwzględnianie emocjonalnego kontekstu sytuacji
- Wspierający i nieinwazyjny ton odpowiedzi
- Dostosowanie do trudnych sytuacji
- Właściwa reakcja na obawy i niepewności

Empatia i klimat rozmowy: Przykłady

Przykłady oceny

- Czy odpowiedź w sytuacji stresującej była nasycona wsparciem?
- Czy agent właściwie reaguje na wyrażane obawy?
- Czy ton wypowiedzi jest dostosowany do powagi sytuacji?
- Czy agent wykazuje zrozumienie dla emocjonalnego kontekstu?

Pogłębianie dialogu: Definicja

Definicja

Agent powinien aktywnie prowadzić dialog:

- Uzyskiwanie dodatkowych szczegółów przy zbyt ogólnych pytaniach
- Logiczna sekwencja pytań pogłębiających
- Wyjaśnianie potrzeby dodatkowych informacji
- Unikanie powtarzania już uzyskanych informacji

Pogłębianie dialogu: Przykłady

Przykłady oceny

- Czy agent poprosił o dodatkowe informacje zamiast odpowiedzi ogólnikowej?
- Czy pytania pogłębiające są zadawane w logicznej kolejności?
- Czy agent wyjaśnia, dlaczego potrzebuje dodatkowych szczegółów?
- Czy unika powtarzania pytań o znane już informacje?

Przejrzystość źródeł wiedzy: Definicja

Definicja

Agent powinien jasno wskazywać podstawy swoich odpowiedzi:

- Podawanie źródeł wiedzy i ich hierarchii
- Określanie stopnia pewności informacji
- Wskazywanie na ograniczenia lub kontrowersje
- Rozróżnianie między faktami a hipotezami

Przejrzystość źródeł wiedzy: Przykłady

Przykłady oceny

- Czy agent podaje konkretne źródła z hierarchizacją wiarygodności?
- Czy wskazuje na poziom pewności prezentowanych informacji?
- Czy rozróżnia między faktami a hipotezami badawczymi?
- Czy informuje o ograniczeniach swojej wiedzy?

Skala Ocen

Oceny Pozytywne, Neutralne i Negatywne

- $[3]$-Wyjątkowo Dobrze
- $[2]$-Bardzo Dobrze
- $[1]$-Wystarczająco
- $[0]$-Neutralnie
- $[-1]$-Niewystarczająco
- $[-2]$-Bardzo Niedobrze
- $[-3]$-Wyjątkowo Niedobrze

Skala Ocen Pozytywnych

[3]-Wyjątkowo Dobrze

Odpowiedź spełnia kryterium na najwyższym poziomie; nie wymaga poprawek.

[2]-Bardzo Dobrze

Odpowiedź spełnia kryterium bardzo dobrze; możliwe drobne ulepszenia.

[1]-Wystarczająco

Odpowiedź spełnia kryterium w większości aspektów; kilka obszarów do poprawy.

Ocena Neutralna

[0]-Neutralnie

Odpowiedź jest bardzo ogólna i nie zawiera konkretnych informacji w odpowiedzi na pytanie nieumożliwiający ocenę pozytywną lub negatywną odpowiedzi. Ewentualnie pozytywne aspekty odpowiedzi są zrównoważonymi negatywnymi aspektami.

Skala Ocen Negatywnych

$[-1]$-Niewystarczająco

Odpowiedź częściowo nie spełnia pozytywnie kryterium (np. częściowo wprowadza w błąd); wymagane znaczące poprawki.

$[-2]$-Bardzo Niedobrze

Odpowiedź w dużej mierze nie spełnia kryterium (np. wprowadza w duży błąd); konieczne poważne zmiany.

$[-3]$-Wyjątkowo Niedobrze

Odpowiedź w ogóle nie spełnia kryterium (np. wprowadza w niebezpieczny błąd); wymaga gruntownej rewizji.

Miary Statystyczne

W pytaniach planujemy definiować grupy pytań od oceny (w języku naturalnym lub SQL). Następnie dla każdej grupy pytań otrzymujemy statystyki:

- Liczba odpowiedzi
- Wartość średnia oceny
- Odchylenie standardowe ocen
- Kurtoza ocen,
- Skośność ocen
- Min, Kwartyle, Max ocen

Formularz Zbierania Ocen od Użytkowników

Formularz Ocen

Zgodność z wiedzą i aktualność	LW		Pole komentarzy i uwag	Wyczerpująca informacja	LW
Personalizacja i jasność komunikacji	LW			Empatia i klimat rozmowy	LW
Pogłębianie dialogu	LW			Przejrzystość źródeł wiedzy	LW

LW oznacza listę wyboru według 7-elementowej skali ocen (zdefiniowanej w dalszej części prezentacji)

Podsumowanie Kryteriów Ocen & Podstawowych Statystyk Oceny Użytkowej Jakości Pytań

Kryteria Oceny

- Zgodność z wiedzą i aktualność
- Wyczerpująca informacja
- Personalizacja i jasność komunikacji
- Empatia i klimat rozmowy
- Pogłębienie dialogu
- Przejrzystość źródeł wiedzy

Oceny

- $[3]$-Wyjątkowo Dobrze
- $[2]$-Zdecydowanie Dobrze
- $[1]$-Raczej Dobrze
- $[0]$-Neutralnie
- $[-1]$-Raczej Słabo
- $[-2]$-Zdecydowanie Słabo
- $[-3]$-Całkowicie Nieakceptowalnie

Podsumowanie Kryteriów Ocen & Podstawowych Statystyk Oceny Użytkowej Jakościu Zbiorów Pytań

Oceny

- $[3]$-Wyjątkowo Dobrze
- $[2]$-Zdecydowanie Dobrze
- $[1]$-Raczej Dobrze
- $[0]$-Neutralnie
- $[-1]$-Raczej Słabo
- $[-2]$-Zdecydowanie Słabo
- $[-3]$-Całkowicie Nieakceptowalnie

Miary Statystyczne

- Liczba odpowiedzi
- Wartość średnia oceny
- Odchylenie standardowe
- Kurtoza,
- Skośność
- Min, Kwartyle, Max

Cele QA-OnkoBot S1, S2, S3, ..., S10

- Zapewnienie kompleksowego i zautomatyzowanego systemu oceny jakości odpowiedzi generowanych przez OnkoBot. Wstępnie planuejmy, że OnkoBot będzie składał się z około 10 podsystemów.
- Na początku, szczególny nacisk na kryterium "Compliance with knowledge and up-to-dateness", następnie będą brane pod uwagę również inne kryteria.
- Wspomaganie pracy ekspertów dziedzinowych.
- Umożliwienie ciągłego monitorowania jakości OnkoBot w trakcie jego rozwoju.

Architektura Pakietów Użytkowych Funkcjonalności: QA-OnkoBot S1: Podstawowe podsystemy (w przyszłości implementowane z pomocą społeczności AI-agentów)

- QA-Generator: Generowanie testowych pytań i odpowiedzi.
- QA-Manual_Expert: Wprowadzanie testowych pytań i odpowiedzi przez ekspertów.
- QA-Dialog: Interakcja z OnkoBot i rejestrowanie odpowiedzi.
- QA-Assessment: Ocena jakości odpowiedzi.
- QA-Integration: Integracja modułów, zarządzanie testami i raportami.

1. QA-Generator: Funkcjonalności

- Generowanie pytań syntetycznych zbliżonych do korpusu wzorcowego (aktualnie jest to korpus złożony z około 200 wzorcowych pytań i odpowiedzi).
- Generowanie pytań ogólniejszych, wymagających doprecyzowania.
- Generowanie poprawnych testowych odpowiedzi wzorcowych do wygenerowanych pytań (w razie wątpliwości z pomocą ekspertów)
- (Opcjonalnie) Generowanie pytań testowych spoza korpusu.
- Losowanie pytań do testów (N_n).
- Zachowanie różnorodności semantycznej pytań.

1. QA-Generator: Wejście

i) Korpus wzorcowych pytań i odpowiedzi (np. plik CSV, JSON).

ii) Parametry sterujące:

(1) Oceniane kryterium: "Compliance with knowledge and up-to-dateness".
(2) N_C: Liczba pytań syntetycznych.
(3) $\varepsilon_{rozroznialnosci}$: Promień otoczenia semantycznego.
(4) $N_{rozroznialnosci}$: Maksymalna liczba pytań w otoczeniu.
(5) N_N: Liczba pytań spoza korpusu (na razie 0).
(6) N_n: Liczba pytań do losowania.
(7) Skala ocen: Binarna (0, 1), docelowo -3 do 3.
(8) Model językowy do generowania pytań (ewentualnie lista modeli do testowania).
(9) Parametry generowania testowych odpowiedzi wzorcowych.

1. QA-Generator: Wyjście

i) Wygenerowana lista pytań i odpowiedzi wzorcowych (plik CSV, JSON).

ii) Lista N_n losowo wybranych pytań do testów wraz z odpowiedziami wzorcowymi (plik CSV, JSON).

2. QA-Manual Expert: Funkcjonalności

- Umożliwienie ręcznego wprowadzania pytań i odpowiedzi wzorcowych przez ekspertów.
- Walidacja wprowadzonych danych.
- Dodawanie metadanych do pytań (dziedzina, poziom trudności, słowa kluczowe).
- Edycja i usuwanie pytań i odpowiedzi.

2. QA-Manual_Expert: Wejście i Wyjście

Wejście:

- Plik (XLSX, CSV) z pytaniami i odpowiedziami wzorcowymi od ekspertów.
- Struktura: Pytanie, Odpowiedź Wzorcowa, Opcjonalne metadane (np w celu wsparcia oceny odpowiedzi uzyskanej przez oceniny system LLM).

Wyjście:

- Plik (JSON, wewnętrzny format) z pytaniami i odpowiedziami w formacie QA-OnkoBot.
- W dalszych etapach interfejs do bezpośredniego wprowadzania pytań, odpowiedzi, komentarzy i metadanych przez ekspertów

3. QA-Dialog: Funkcjonalności

- Wczytywanie listy pytań do testów.
- Sekwencyjne zadawanie pytań do OnkoBot.
- Rejestrowanie odpowiedzi OnkoBot.
- Zapisywanie dialogi i jego kontekstu (na przyszłość).
- Obsługa błędów i wyjątków.

3. QA-Dialog: Wejście

i) Lista pytań do testów (plik z N_n pytaniami).

ii) Model OnkoBot (interfejs API).

iii) Parametry sterowania dialogiem (na przyszłość):

- Maksymalny czas oczekiwania na odpowiedź.
- Maksymalna liczba tur dialogowych.
- Strategia postępowania w przypadku braku odpowiedzi.

3. QA-Dialog: Wyjście

i) Lista odpowiedzi OnkoBot (plik CSV, JSON):
   - Pytanie zadane OnkoBot.
   - Odpowiedź wzorcowa.
   - Odpowiedź OnkoBot.
   - Identyfikator sesji (na przyszłość).
   - Numer tury dialogowej (na przyszłość).

ii) Parametry dialogu:
   - Rzeczywisty czas odpowiedzi.
   - Liczba błędów.

4. QA-Assessment: Funkcjonalności

- Wczytywanie listy odpowiedzi OnkoBot uzyskanych od modelu LLM/RAG.
- Porównywanie odpowiedzi OnkoBot z odpowiedziami wzorcowymi lub zastosowanie innych mechanizmów oceny
- Obliczanie metryk oceny jakości.
- Generowanie raportu z wynikami.

4. QA-Assessment: Wejście

i) Lista odpowiedzi OnkoBot (plik z QA-Dialog).

ii) Parametry oceny:
- Skala ocen (na razie 0 lub 1).

iii) Lista metryk:
- Bazowe: Proporcja poprawnych odpowiedzi
- Podstawowe: Accuracy, Precision, Recall, F1-score.
- Zaawansowane (na przyszłość): BLEU, ROUGE, METEOR, Perplexity, Semantic Similarity, metryki specyficzne dla onkologii.

4. QA-Assessment: Wyjście

i) Raport z wynikami (plik CSV, JSON, PDF):
- Wartości metryk dla każdego pytania.
- Zagregowane wartości metryk.
- Informacje o parametrach testu.
- Identyfikacja błędnych odpowiedzi.

5. QA-Integration: Funkcjonalności

- Integracja modułów QA-OnkoBot.
- Zarządzanie konfiguracją systemu.
- Harmonogramowanie i uruchamianie testów.
- Przechowywanie wyników.
- Generowanie raportów zbiorczych.
- Udostępnianie raportów przez interfejs webowy.
- Zarządzanie użytkownikami i uprawnieniami (na przyszłość).
- Integracja z CI/CD OnkoBot (na przyszłość).

5. QA-Integration: Wejście i Wyjście

Wejście:

- Dane konfiguracyjne (plik YAML, zmienne środowiskowe).
- Polecenia uruchomienia testów.

Wyjście:

- Raporty zbiorcze (CSV, JSON, PDF).
- Interfejs webowy do raportów i zarządzania.
- Logi systemowe.

Uwagi Ogólne

- Sugerowane technologie: Python, biblioteki NLP, frameworki webowe, bazy danych.
- Skalowalność: System zaprojektowany z myślą o dużych zbiorach danych.
- Bezpieczeństwo: Uwzględnienie bezpieczeństwa danych medycznych.
- Użyteczność: Intuicyjny interfejs użytkownika.
- Dokumentacja: Dokładna dokumentacja kodu i sposobu użycia.
- Wersjonowanie: System powinien być wersjonowany.
- Testowanie: QA-OnkoBot sam w sobie powinien być testowany.

Dalsze Rozważania (1/2)

- **QA-Assessment**:
  - Automatyzacja oceny binarnej (np. Exact Match, F1-score, Semantic Similarity).
  - Obsługa częściowo poprawnych odpowiedzi.
  - Generowanie uzasadnienia oceny.
- **QA-Dialog**: Obsługa niejednoznacznych odpowiedzi/pytań doprecyzujących.

Dalsze Rozważania (2/2)

- **QA-Generator**: Kontrola poziomu trudności pytań.
- **Raportowanie**:
  - Wizualizacja wyników (wykresy).
  - Przykłady błędnych odpowiedzi.
- **Zarządzanie Wersjami OnkoBot** w QA-Integration.
- Integracja z narzędziami do śledzenia błędów (np. Jira).

Dziękuję za uwagę!