

RADA NAUKOWA DYSCYPLINY
INFORMATYKA TECHNICZNA I TELEKOMUNIKACJA POLITECHNIKI WARSZAWSKIEJ

zaprasza na
OBRONĘ ROZPRAWY DOKTORSKIEJ

mgr inż. Karoliny Teresy Gabor-Siatkowskiej

która odbędzie się w dniu 16.03.2026 roku, o godzinie 12.00 w trybie stacjonarnym.

Temat rozprawy: Improving Therapeutic Spoken Dialogue Systems with Eye Tracking

Promotor: dr hab. inż. Artur Janicki, prof. uczelni – Politechnika Warszawska

Recenzenci: dr hab. inż. Katarzyna Hareźniak, prof. uczelni, Wydział Automatyki, Elektroniki i Informatyki, Politechnika Śląska
dr hab. inż. Tomasz Kapuściński, prof. uczelni, Wydział Elektrotechniki i Informatyki, Politechnika Rzeszowska
dr hab. inż. Aleksandra Kawala-Sterniuk, prof. uczelni, Wydział Informatyki i Telekomunikacji, Politechnika Wrocławska

Obrona odbędzie się w sali nr 478 w Gmachu Wydziału Elektroniki i Techniki Informacyjnych Politechniki Warszawskiej. Z rozprawą doktorską i recenzjami można zapoznać się w Czytelni Biblioteki Głównej Politechniki Warszawskiej, Plac Politechniki 1, oraz na stronie internetowej BIP PW:

<https://www.bip.pw.edu.pl/Postepowania-w-sprawie-nadania-stopnia-naukowego/Doktoraty/Wszczete-po-30-kwietnia-2019-r/Rada-Naukowa-Dyscypliny-Informatyka-Techniczna-i-Telekomunikacja/mgr-inz.-Karolina-Gabor-Siatkowska>

Przewodniczący Rady Naukowej Dyscypliny
Informatyka Techniczna i Telekomunikacja
Politechniki Warszawskiej
Prof. dr hab. inż. Jarosław Arabas

Streszczenie

Tematem niniejszej rozprawy jest wykorzystanie technologii śledzenia ruchu gałek ocznych (okulografii) w celu udoskonalenia głosowego systemu dialogowego do celów terapeutycznych. System ten został stworzony na potrzeby pacjentów z rozpoznaniem schizofrenii. Podczas rozmów pacjentów z Terabotem, wystąpił szereg problemów, głównie związanych z płynnością dialogu. W celu znalezienia odpowiednich rozwiązań, wykorzystano dane okulograficzne.

Obecnie okulografia jest szeroko wykorzystywana zarówno w badaniach naukowych jak i zastosowaniach komercyjnych obejmujących różnorodne dziedziny, np. medycynę, edukację, psychologię czy marketing. Przykładem zastosowania w dziedzinie "komunikacji człowiek – komputer" są tzw. agenty konwersacyjne lub systemy dialogowe, którym poświęca się ostatnio wiele uwagi. Jednym z zastosowań jest np. wykorzystanie zachowań niewerbalnych do nawiązania kontaktu z użytkownikiem. Kluczową koncepcją w projektowaniu tych systemów jest wykorzystanie zachowań niewerbalnych do zbierania informacji o użytkownikach, zamiast polegania wyłącznie na mowie. Aby dialog między agentem a człowiekiem stał się bardziej realistyczny, agenci mogą być wyposażeni w zdolności komunikacyjne i ekspresyjne podobne do tych obserwowanych w interakcji z ludźmi, w czasie której mamy do czynienia z ich mową, gestami mimiką twarzy czy wzrokiem). Dzięki temu dialog może przebiegać bardziej naturalnie, a systemy mogą zapewnić bardziej realistyczną rozmowę z użytkownikami. Dane ze śledzenia ruchu gałek ocznych mogą służyć jako dodatkowe źródło informacji, szczególnie podczas rozmów z agentami dialogowymi w aspekcie zdrowia psychicznego.

Terapeutyczny system dialogowy Terabot został zaprojektowany przez zespół badawczy z Politechniki Warszawskiej (PW), Wydziału Elektroniki i Technik Informacyjnych. Jego utworzenie miało na celu wzbogacenie terapii pacjentów z zaburzeniami takimi jak np. depresja, zaburzenia lękowe itp. W celu zweryfikowania skuteczności systemu przeprowadzono pilotażowe randomizowane badanie kliniczne w Instytucie Psychiatrii i Neurologii w Warszawie. 38 pacjentów miało możliwość przeprowadzania rozmów terapeutycznych z systemem dialogowym Terabot (oprócz bieżącego leczenia, w tym farmakologicznego). Podczas rozmowy z Terabotem pacjent mógł wybrać jedną z trzech emocji: lęk, strach lub wstyd. Pod koniec rozmowy Terabot zachęcał do wykonania ćwiczenia relaksacyjnego, które miało pomóc pacjentom w uspokojeniu się i opanowaniu emocji. Zbierane były dane okulograficzne oraz nagrania audio z dialogów pomiędzy pacjentami a Terabotem. Bezpośredni kontakt z pacjentami pozwolił na znacznie efektywniejszą optymalizację systemu dialogowego.

Na podstawie analiz zweryfikowano dotychczasowe działanie istniejącego systemu dialogowego. Przeprowadzono analizy interfejsu graficznego Terabota, w szczególności obszarów zainteresowania pacjentów (ang. area of interest, AOI). Analizowano gdzie pacjenci patrzyli najdłużej i najczęściej podczas gdy kończyli swoją wypowiedź. Wyniki analiz stały się kluczowym źródłem informacji, które posłużyły do zaprojektowania kolejnej wersji systemu. Nowy system jest systemem multimodalnym, wspomaganym przez sygnał okulograficzny. Jest on

w stanie reagować na zachowanie pacjenta (jego spojrzenie) podczas rozmowy. Ponadto system ten opiera się na dużych modelach językowych (ang. Large Language Models, LLM), w szczególności na jednym z modeli GPT firmy OpenAI. Takie podejście sprawia, że nowy system jest bardziej zaawansowany nie tylko naukowo, ale i technologicznie. Dzięki zastosowaniu okulografu (w przeciwieństwie do systemu dialogowego bez dodatkowych narzędzi) można było dostosować konwersacje do zachowania pacjenta. Pozwalało to na bardziej naturalną interakcję między pacjentem a systemem dialogowym. Dane okulograficzne posłużyły do rozwiązania problemów, które powodowały przerwy w dialogu między systemem a pacjentem. Ponadto technologię śledzenia ruchu gałek ocznych wykorzystano również do automatycznej oceny zaangażowania pacjentów w dialog. Proponowane rozwiązania mogą mieć pozytywny wpływ na interakcję pomiędzy Terabotem a pacjentem. Może się to również przyczynić do zwiększenia dobrego samopoczucia pacjentów i komfortu podczas sesji terapeutycznych.

Słowa kluczowe: okulografia, system dialogowy, interakcja człowiek – komputer, systemy multimodalne, psychiatria, komputerowe wspomaganie medycyny, terapia poznawczo - behawioralna, fiksacje, duże modele językowe, GPT, automatyczna ocena zaangażowania, kalibracja

Dr hab. inż. Katarzyna Haręźlak, prof. PŚ
Katedra Informatyki Stosowanej
Wydział Automatyki, Elektroniki i Informatyki
Politechnika Śląska w Gliwicach

Recenzja rozprawy doktorskiej

Pani mgr inż. Karoliny Gabor-Siatkowskiej
pt. „Improving Therapeutic Spoken Dialogue Systems with Eye Tracking”

przygotowanej pod kierunkiem
dr. hab. inż. Artura Janickiego, prof. PW

Niniejsza recenzja została przygotowana na podstawie uchwały Rady Naukowej Instytutu Informatyki Politechniki Warszawskiej z dnia 23.09.2025 w przedmiocie wyznaczenia recenzentów niniejszej pracy doktorskiej (Uchwała nr 81/205).

1. Przedmiot rozprawy, aktualność i ważność tematyki rozprawy

Aktualne badania wskazują, że zdrowie psychiczne stało się globalnym wyzwaniem, a dostęp do terapii jest ograniczony przez niedobór specjalistów oraz związane z nią koszty. Technologie cyfrowe, w tym duże modele językowe (LLM) i roboty terapeutyczne, stają się więc kluczowym elementem odpowiedzi na rosnący kryzys zdrowia psychicznego. Systemy LLM umożliwiają naturalną komunikację z pacjentem, personalizację ćwiczeń oraz adaptację scenariuszy terapii w czasie rzeczywistym — co jest szczególnie istotne w przypadku pacjentów, którzy mają trudności w kontakcie z terapeutą lub odczuwają wysoki poziom stresu. Robot terapeutyczny jest bezstronny, nie ocenia i zapewnia stałą, spokojną reakcję, co ułatwia budowanie poczucia bezpieczeństwa i zaufania pacjenta.

Rozwiązania tego typu nie zastępują jednak terapeuty, lecz stanowią dla niego wsparcie oraz zwiększają dostępność terapii w miejscach, gdzie brakuje specjalistów. Mogą one dodatkowo rejestrować szczegółowe dane o postępach i przebiegu terapii, co daje terapeutom nowe narzędzia do analizy, personalizacji programu terapeutycznego i obiektywnej oceny jej efektów. Praca takiego robota może być wspomagana różnymi metodami, jak pomiar biometrycznych danych fizjologicznych, analizę głosu czy tonu mowy. Dzięki dynamicznie rozwijającej się dziedzinie okulografii i metodom śledzenia ruchu oczu, możliwa stała się również ocena uwagi i koncentracji pacjenta oraz częstotliwości kontaktu wzrokowego. Taka analiza może dostarczyć pełniejszej oceny stanu emocjonalnego pacjenta, zapewniając bardziej precyzyjny zestaw reakcji i lepsze efekty terapii. Dodatkowo, zapis ruchu oka może stanowić uzupełnienie mechanizmu uwagi modelu LLM, co w konsekwencji może poprawić dobór słów i jakość konwersacji.

Biorąc powyższe pod uwagę, należy stwierdzić, że rozpatrywane w pracy problemy są aktualne i dobrze wpisują się w nurt obecnie prowadzonych prac naukowych. Tematykę podjętą w rozprawie doktorskiej przez Panią magister inżynier Karolinę Gabor-Siatkowską, należy uznać za ważną, celową

i aktualną. Niewątpliwie tematyka ta wpisuje się w dyscyplinę Informatyka Techniczna i Telekomunikacja.

W rozprawie postawiono jedną, dość ogólną hipotezę:

Wykorzystanie śledzenia ruchu gałek ocznych może zwiększyć efektywność terapeutycznego systemu dialogowego opartego na mowie (ang. The use of eye tracking can improve a therapeutic spoken dialogue system).

Jej uszczegółowienia można szukać w postawionych celach badawczych, wśród których znalazły się:

1. Zbadanie, czy konieczna jest kalibracja okulografu podczas jego użycia w systemie dialogowym opartym na mowie (ang. Investigating whether an eye tracker needs to be calibrated when used in a spoken dialogue system).
2. Analiza obszarów zainteresowania w interfejsie graficznym systemu dialogowego (ang. Analysis of areas of interest in the graphical interface of the dialogue system).
3. Zwiększenie płynności rozmów z systemem dialogowym poprzez zastosowanie urządzenia do śledzenia ruchu gałek ocznych (ang. Enhancement of fluency of conversations with a dialogue system by using an eye tracker).
4. Umożliwienie automatycznej oceny zaangażowania użytkownika podczas interakcji z systemem dialogowym (ang. Enabling automatic user engagement assessment while interacting with the dialogue system).

2. Zawartość rozprawy

Rozprawa doktorska składa się z ośmiu rozdziałów i dwóch dodatków. W rozdziale pierwszym zawarto wprowadzenie w tematykę rozprawy, przedstawiono motywację prowadzonych badań oraz cele rozprawy. W rozdziale tym podsumowano również aktywność badawczą i akademicką Doktorantki, jak również uzyskane przez Nią nagrody. W rozdziale drugim, na podstawie przeglądu literaturowego, dokonano oceny aktualnego stanu wiedzy w dziedzinie rozprawy. W tej części rozprawy znalazła się analiza najczęściej występujących schorzeń psychicznych i ich wpływu na życie społeczeństw, ze szczególnym uwzględnieniem sytuacji w Polsce. Omówiono też istniejące rozwiązania w zakresie systemów dialogowych, kierunki ich rozwoju oraz podstawową architekturę systemu dialogu mówionego. Następnie zaprezentowano konkretne rozwiązania w dziedzinie ochrony zdrowia psychicznego. Ostatnia część tego rozdziału poświęcona została technologii śledzenia ruchu oka, jej zastosowaniom w interakcji człowiek – komputer oraz w dziedzinie rozprawy doktorskiej. W rozdziale trzecim zawarto prezentację terapeutycznego systemu dialogowego Terabot oraz eksperymentów zrealizowanych z jego wykorzystaniem. Wyniki przeprowadzonych badań pozwoliły na wskazanie problemów utrudniających przeprowadzenie płynnej konwersacji terapeutycznej oraz posłużyły do określenia celów rozprawy. Kolejne rozdziały poświęcono realizacji postawionych celów. Każdy z nich rozpoczyna się uzasadnieniem dla podjętych badań, następnie przedstawia ich wykonanie oraz analizę i podsumowanie wyników. W rozdziale czwartym zaprezentowano opis oraz wyniki badań dotyczących oceny konieczności przeprowadzenia kalibracji przed właściwym zastosowaniem okulografu. W eksperymencie poprawnie zastosowano wszystkie procedury niezbędne do uzyskania rzetelnych wyników. W rozdziale piątym omówiono metodę wyznaczania wzorców wzrokowych uczestników spotkań terapeutycznych, uwzględniającą zapisy ruchu oka oraz obszary zainteresowań (AOI) określone w interfejsie systemu Terbot. Rozdział ten zawiera również analizę wyników uzyskanych po

zastosowaniu zaproponowanej metody do zapisów ruchu oka pacjentów z rozpoznaniem zaburzeń psychicznych. W rozdziale szóstym zaprezentowano nową wersję systemu terapeutycznego, w której do prowadzenia dialogu użyto system językowy LLM, a dla zapewnienia płynności dialogu zapisy ruchu oka. Przedstawiono w nim metodę aktywacji systemu terapeutycznego dla przypadków przedłużającego się czasu oczekiwania na wypowiedź pacjenta. Zaproponowany został także próg czasowy aktywacji, wyznaczony na podstawie zapisów z wcześniej zarejestrowanych spotkań terapeutycznych. W tej części rozprawy opisane zostały również eksperymenty weryfikujące działanie metody w warunkach zbliżonych do rzeczywistych, z udziałem profesjonalnych aktorów odgrywających rolę pacjentów. Wyniki eksperymentów, zestawione z wynikami uzyskanymi dla poprzedniej wersji systemu, potwierdziły większą skuteczność nowej wersji systemu Terabot w utrzymywaniu płynności dialogu. Zdiagnozowano również nieprawidłowe działania systemu i ich przyczyny. W rozdziale siódmym, uwagę poświęcono zagadnieniom automatycznej oceny obecności i zaangażowania pacjenta w trakcie rozmowy terapeutycznej. Wykazano w nim brak spójności opisów przebiegu takiej rozmowy, przygotowanych przez różnych asystentów, co skłoniło Doktorantkę do zaproponowania metryk pozwalających na bardziej obiektywną ocenę. Bazują one na transkrypcji dialogu oraz analizie ruchu oka. Ostatni rozdział rozprawy zawiera podsumowanie osiągnięć, dyskusję ograniczeń i możliwych kierunków dalszych badań w obszarze rozprawy doktorskiej. Na końcu rozprawy zamieszczono dwa dodatki obejmujące przykładowe scenariusze dialogów z systemem Terabot i zadań eksperymentalnych.

Podsumowując, można stwierdzić, że rozprawa została skonstruowana poprawnie i zawiera wszystkie oczekiwane treści. Można jednak mieć pewien niedosyt związany z technicznym opisem zaproponowanych rozwiązań. W rozprawie doktorskiej przygotowywanej w dyscyplinie Informatyka Techniczna i Telekomunikacja zasadne byłoby zamieszczenie bardziej formalnego opisu użytych algorytmów i metod.

3. Analiza źródeł

Autorka rozprawy wykazała się dobrą znajomością dorobku literaturowego dotyczącego zagadnień, którym poświęcona jest rozprawa. Podstawą oceny literatury jest spis liczący 140 pozycji (siedem pozycji jest współautorstwa Doktorantki), związanych z tematyką rozprawy. W większości są to źródła recenzowane, niekiedy źródła internetowe, opublikowane w ciągu ostatnich dziesięciu lat. Są one cytowane w odpowiednim kontekście. Choć w samym spisie widoczne są drobne uchybienia, takie jak brak autorów w pozycjach [116], [122] oraz nie do końca trafnie umieszczone odwołanie do pozycji [139] w podsumowaniu rozprawy, dobór literatury oceniam pozytywnie. Świadczy on o dobrym rozeznaniu Doktorantki w literaturze światowej w tematyce, którą się zajmuje.

4. Oryginalne osiągnięcia

Do najważniejszych osiągnięć rozprawy, decydujących o jej wartości teoretycznej i praktycznej, a które wyróżniają ją spośród dostępnych w literaturze przedmiotu, zaliczyć można:

- a. opracowanie eksperymentu i badania dotyczące konieczności kalibracji okulo grafu w celu pomiaru średnicy źrenicy,
- b. opracowanie metody aktywacji systemu Terabot na podstawie analizy ruchu oka, co objęło:

- wyznaczenie fazy zamykającej ostatniego zdania wypowiedzi,
 - wyznaczenie obszarów zainteresowań (AOI) istotnych z punktu widzenia rozwiązywanego problemu,
 - określenie sposobu wykorzystania wyników metody w rzeczywistym środowisku terapeutycznym,
- c. opracowanie nowej wersji systemu dialogowego bazującej na dużym modelu językowym, LLM, oraz danych ruchu oka, a w tym:
- dobór parametrów modelu LLM zapewniających charakterystykę wypowiedzi systemu Terabot właściwą dla celów terapeutycznych,
 - zapewnienie aktualnych i zweryfikowanych danych dzięki zastosowaniu modelu generatywnego z wyszukiwaniem informacji w zewnętrznej bazie wiedzy (technika RAG),
 - przeprowadzenie dwuetapowej weryfikacji zaproponowanej metody: na podstawie zapisów ruchu oka osób ze stwierdzonymi zaburzeniami psychicznymi oraz w trakcie sesji terapeutycznej z uczestnikami odtwarzającymi zachowania osób z zaburzeniami,
- d. opracowanie metody automatycznie oceniającej zaangażowanie pacjenta, bazującej na transkrypcji dialogu oraz analizie ruchu oka.

Podsumowując należy zauważyć, że zaproponowane rozwiązania w postaci metod usprawniających przebieg spotkania terapeutycznego z zastosowaniem systemu Terabot i z uwzględnieniem analizy ruchu oka, poza wymiarem naukowym, mają również wymiar praktyczny: mogą stanowić wsparcie w pracy terapeutów oraz łagodzić trudności wynikające z ich ograniczonej dostępności.

Potwierdzeniem osiągnięć Doktorantki jest wskazany w rozprawie dorobek publikacyjny, na który składa się 10 publikacji naukowych, z czego dwie to pozycje z wyznaczonym współczynnikiem wpływu (IF) o sumarycznej wartości 5,029. Obie publikacje znajdują się na liście ministerialnej z liczbą 140 i 100 pkt. MNiSW (w rozprawie błędnie podano punktację dla publikacji oznaczonej numerem 6). Jedna z tych publikacji nie jest związana z dziedziną rozprawy. Na pozostałe publikacje składają się: trzy artykuły w czasopismach (70, 20 i 5 pkt. MNiSW), dwa materiały konferencyjne (140 i 20 pkt. MNiSW), oraz trzy rozdziały w monografii (każdy 20 pkt. MNiSW). Wszystkie prace są wieloautorskie. W ośmiu z nich Doktorantka jest pierwszą autorką, włączając w to publikację z wyznaczonym współczynnikiem wpływu (IF: 2,6 i 100 pkt. MNiSW). Sześć, ze wskazanych publikacji, indeksowanych jest w bazie Scopus, a dwie w bazie Web of Science. W **bazie Scopus Indeks Hirscha** Doktorantki wynosi **4**, a w **Web of Science** **2**. Ponadto Doktorantka brała udział w pięciu projektach badawczych powiązanych zakresem z tematyką rozprawy oraz w seminariach organizowanych w ramach tych projektów, prezentując wyniki swoich prac.

5. Wady i słabe strony rozprawy, uwagi dyskusyjne

Rozprawa jako całość nie ma istotnych wad w zakresie metody badawczej, sposobu prezentacji wyników, a także w zakresie formułowania wniosków. Cele pracy zostały prawie w całości zrealizowane, a postawiona ogólna hipoteza została potwierdzona. Jednak rozprawa posiada również słabe strony. Do najpoważniejszych z nich należą brak formalnego zapisu algorytmicznego zaproponowanych metod oraz ograniczona liczba uczestników badań. Zapewnienie odpowiedniej liczby uczestników stanowi powszechnie znane wyzwanie w badaniach wymagających udziału osób,

szczególnie tych, z problemami zdrowotnymi. Zachęcam jednak do przeprowadzenia kolejnych eksperymentów, które zapewnią silniejsze potwierdzenie uzyskanych w rozprawie wyników.

Ponadto w tekście rozprawy są pewne zagadnienia, które wymagają dodatkowego wyjaśnienia:

- a. Na stronie 40, wymieniono parametry różnych zdarzeń ruchu oka, część z nich wydaje się nieprecyzyjna. Dla przykładu, trudny do wyobrażenia dla zdrowej osoby jest czas opóźnienia sakkadowego wynoszącego 1000 ms. Tak długie czasy opóźnienia traktuje się raczej jako sytuacje patologiczne. Podobnie dla fiksacji, typowy zakres to 150-300 ms. Natomiast, zakres czasu dla sakkad między 20 a 40 ms wydaje się zawężony. Czas ten zależy od amplitudy ruchu i może u zdrowej osoby sięgać nawet 80-100ms.
- b. W tabeli 4, na stronie 53, wśród parametrów okulo grafu używanego w badaniach w Instytucie Psychiatrii i Neurologii, wymieniona jest kalibracja pięciopunktowa. Z treści rozprawy nie wynika, czy była ona przeprowadzona przed rozpoczęciem dialogu terapeutycznego. A jeśli tak, jaki błąd kalibracji był akceptowany/dopuszczający pacjenta do badania? Czy taka kalibracja stanowiła problem dla pacjentów?
- c. Na stronie 57, pojawia się zdanie: *Pacjenci czuli również, że ich odpowiedzi i emocje zostały zrozumiane (3,5 w skali Likerta)*. Rysunek 27 nie potwierdza tego wniosku, szczególnie w porównaniu z elementem umieszczonym wyżej na wykresie.
- d. W Tabeli 7, na stronie 83, zaprezentowano porównanie liczby płynnie przeprowadzonych konwersacji terapeutycznych z użyciem starej i nowej wersji systemu Terabot. Ocena tych wyników nie jest jednak możliwa, gdyż w pierwszym wierszu tabeli znajdują się błędy liczbowe: wartości nie sumują się poprawnie.
- e. Rysunek 46, na stronie 94, zawiera propozycję opisu przebiegu spotkania dla psychiatry. Czy tekst wygenerowany został automatycznie przez jeden z modułów systemu, czy jest to propozycja Doktorantki? Jeśli jest to praca Doktorantki, to czy można zautomatyzować przygotowanie takiego opisu na podstawie zaproponowanej macierzy oceny zaangażowania pacjenta?

Uwagi o charakterze dyskusyjnym

1. Badania dotyczące pomiaru średnicy źrenicy zdają się tylko częściowo pokrywać pierwszy cel rozprawy, zdefiniowany jako: *Zbadanie czy śledzenie wzroku wymaga kalibracji w przypadku użycia systemu bazującym na dialogu mówionym*. Pozyskiwanie informacji o konkretnym punkcie lub obszarze spojrzenia pacjenta, bez przeprowadzonej kalibracji, może być utrudnione i wymaga specjalizowanych podejść. Przeprowadzone eksperymenty nie rozwiązują tego problemu.
2. W rozdziale piątym wspomniano, że synchronizacja mowy i ruchu oka uzyskiwana jest dzięki funkcjonalności wykorzystywanego okulo grafu. W tym kontekście należy zapytać:
 - o działania, które musiałyby być podjęte w przypadku użycia innego urządzenia,
 - oraz czy synchronizowane fiksacje (FPOGX, FPOGY) zostały pozyskane z wcześniejszą kalibracją czy bez niej?
3. Również w rozdziale piątym zaprezentowano cechy wzorca wzrokowego wyznaczonego podczas końcowych faz wypowiedzi. Brak odniesienia tych wyników do pozostałych części wypowiedzi nie pozwala na stwierdzenie, czy i w jakim stopniu zmieniał się sposób obserwacji interfejsu systemu Terabot we wcześniejszych fazach spotkania.

4. Porównanie poprzedniej i nowej, opracowanej w rozprawie, wersji systemu Terabot dokonano stosując „niesymetryczne” dane wejściowe: dość mocno różniła się liczba porównywanych dialogów oraz jednak trochę inna charakterystyka osób uczestniczących w dialogu – pacjenci i osoby odgrywające rolę pacjentów. Choć wyniki takiego porównania są optymistyczne (bazując na nie do końca poprawnych treściach Tabeli 7), rodzi się pytanie, czy wyrównanie warunków eksperymentu dla nowej wersji systemu - większa liczba dialogów z osobami ze zdiagnozowaną chorobą psychiczną, zachowałaby charakterystykę porównania.
5. Ustawienie wartości progowej wyznaczono na podstawie dość ograniczonej liczby konwersacji. Czy istnieje szansa, że zastosowana wartość progowa uległaby znaczącej zmianie, gdyby badania przeprowadzono na większej próbie danych? Czy wprowadzenie procedury z adaptacyjnym progiem aktywacji systemu, dostosowanego do konkretnego pacjenta, mogłoby zapewnić płynny przebieg z dowolną osobą, uwzględniając charakterystykę sposobu jej wypowiedzania się?

6. Wniosek końcowy

W podsumowaniu stwierdzam, że rozprawa doktorska mgr inż. Karoliny Gabor-Siatkowskiej, mimo pewnych uwag krytycznych, **spełnia wymagania stawiane kandydatom do stopnia naukowego doktora**, określone w art. 187 ust. 1 i 2 Ustawy z dnia 20 lipca 2018r. Prawo o szkolnictwie wyższym i nauce (Dz.U. z 2020r. poz. 85, z późn. zm.). W mojej ocenie zawiera **oryginalne rozwiązanie problemu naukowego**. Doktorantka osiągnęła większość postawionych celów, wykazując się niezbędną wiedzą teoretyczną w dyscyplinie Informatyka Techniczna i Telekomunikacja i umiejętnościami w zakresie prowadzenia badań naukowych z wykorzystaniem metod informatycznych. **Wnoszę zatem o dopuszczenie przedłożonej do recenzji rozprawy do publicznej obrony.**

Dr hab. inż. Katarzyna Haręźlak, prof. PŚ



dr hab. inż. Tomasz Kapuściński, prof. PRz
Katedra Informatyki i Automatyki
Wydział Elektrotechniki i Informatyki
Politechnika Rzeszowska im. Ignacego Łukasiewicza
al. Powstańców Warszawy 12
35-029 Rzeszów

Rzeszów, 5 grudnia 2025 r.

Recenzja

rozprawy doktorskiej

mgr inż. Karoliny Teresy Gabor-Siatkowskiej

pt. *Improving Therapeutic Spoken Dialogue Systems with Eye Tracking*

1 Tematyka rozprawy

Rozprawa doktorska Pani mgr inż. Karoliny Teresy Gabor-Siatkowskiej dotyczy zaawansowanego, multimodalnego systemu interakcji człowiek-komputer wspomagającego terapię osób z problemami psychicznymi. System w postaci głosowego agenta konwersacyjnego powstał na bazie istniejącego rozwiązania poprzez dodanie do kanału wejściowego danych pozyskiwanych z urządzenia do śledzenia ruchów gałek ocznych oraz zamianę algorytmów przetwarzania języka naturalnego na rozwiązanie wykorzystujące duży model językowy. Tezę naukową pracy sformułowano następująco: **Wykorzystanie technologii śledzenia ruchu gałek ocznych może poprawić działanie głosowego systemu dialogowego do celów terapeutycznych.** Zdefiniowano cztery zadania badawcze:

- Sprawdzenie, czy urządzenie do śledzenia ruchu gałek ocznych wymaga kalibracji podczas stosowania w głosowym systemie dialogowym.
- Analiza obszarów zainteresowania w graficznym interfejsie systemu dialogowego.
- Poprawa płynności konwersacji w systemie dialogowym dzięki zastosowaniu urządzenia do śledzenia ruchu gałek ocznych.
- Umożliwienie automatycznej oceny zaangażowania użytkownika podczas interakcji z systemem dialogowym.

2 Charakterystyka rozprawy

Rozprawa składa się z ośmiu rozdziałów i dwóch dodatków.

W rozdziale pierwszym, mającym charakter wprowadzenia, przedstawiono tło podejmowanych prac badawczych, motywację, która przyświecała autorce, cel i zakres pracy oraz krótką charakterystykę dorobku doktorantki.

Rozdział drugi poświęcony jest przeglądowi obecnego stanu wiedzy w trzech kluczowych obszarach tematycznych pracy, obejmujących problematykę zdrowia psychicznego i opieki psychiatrycznej, agentów konwersacyjnych oraz technologię śledzenia ruchu gałek ocznych.

W rozdziale trzecim przedstawiono interfejs użytkownika i architekturę systemu Terabot, dialogowego asystenta w języku polskim dla pacjentów z zaburzeniami psychicznymi. Opisano również problemy z płynnością interakcji, zidentyfikowane podczas badań w Instytucie Psychiatrii i Neurologii w Warszawie, obejmujące przerywanie wypowiedzi, długi czas oczekiwania na odpowiedź pacjenta i agenta oraz brak informacji o zachowaniu pacjenta podczas sesji terapeutycznej.

Potrzebę kalibracji urządzenia śledzącego wzrok w systemie dialogowym przeanalizowano w rozdziale czwartym. Opisano problem, metodykę eksperymentu oraz wyniki, które wskazują, że kalibracja nie jest krytycznym wymogiem dla poprawnego działania systemu.

W rozdziale piątym przeanalizowano dane z śledzenia wzroku pacjentów podczas rozmów z systemem Terabot, opisano metodykę badawczą i uzyskane wyniki. Proponowane kliniczne zastosowanie rozwiązania zakładało aktywację odpowiedzi systemu po zakończeniu wypowiedzi użytkownika, gdy jego wzrok był skupiony na określonym obszarze ekranu. Pomysł ten porzucono jednak ze względu na potencjalne wykluczenie pacjentów, u których utrzymanie kontaktu wzrokowego może być utrudnione.

Projekt i testy terapeutycznego systemu dialogowego opartego na dużym modelu językowym i danych z śledzenia wzroku opisano w rozdziale szóstym. Przeprowadzono analizę czasu oczekiwania pacjentów, a następnie przedstawiono implementację systemu, testy offline i real-time z udziałem pacjentów i aktorów, analizę działania systemu i ocenę satysfakcji użytkowników.

Rozdział siódmy poświęcony jest automatycznej ocenie zaangażowania pacjentów. Opisano problem, dokonano analizy notatek asystentów dokonujących takiej oceny, a następnie przedstawiono i przebadano proponowaną metodę automatyczną wykorzystującą dane z systemu do śledzenia ruchów gałek ocznych.

Rozdział ósmy stanowi podsumowanie całej pracy oraz dyskusję na temat osiągnię-

tych wyników. Dodatkowo, omówiono w nim ograniczenia proponowanego systemu oraz kierunki dalszych badań.

Dodatek A zawiera przykładowy dialog dotyczący uczucia złości, natomiast Dodatek B scenariusze przygotowane dla aktorów i aktorek z Teatru PW, którzy brali udział w testach.

3 Ocena rozprawy

W kontekście zidentyfikowanych w rozprawie potrzeb w obszarze opieki psychiatrycznej, badania doktorantki mają istotne znaczenie społeczne. Podejmowane prace, charakteryzujące się solidnym uzasadnieniem i przekonującym podejściem, stanowią znaczący wkład w rozwój nowej generacji narzędzi terapeutycznych.

Nowatorskie podejście doktorantki polega na integracji technologii śledzenia ruchu gałek ocznych z systemem dialogowym, co ma na celu poprawę płynności interakcji. Brak analogicznych rozwiązań w języku polskim potwierdza oryginalność prowadzonych badań.

Doktorantka osiągnęła znaczący postęp w udoskonalaniu systemu dialogowego, wykorzystując śledzenie ruchu gałek ocznych. Udało się zaproponować rozwiązanie problemów związanych z płynnością dialogu, w tym wykrywanie zakończenia wypowiedzi pacjenta (endpoint detection) oraz ograniczenie problemów związanych z długimi czasami oczekiwania na odpowiedzi zarówno ze strony agenta, jak i pacjenta. Dodatkowo, opracowano mechanizm automatycznej oceny zachowania pacjenta podczas interakcji, a w szczególności podczas ćwiczeń relaksacyjnych, co umożliwiła bardziej precyzyjną i spersonalizowaną interwencję. Prace te zaowocowały zaprojektowaniem ulepszonej wersji systemu dialogowego, uwzględniającej zebrane dane i wykorzystującej potencjał śledzenia ruchu gałek ocznych.

Należy dodać, że badania nad udoskonalaniem systemów dialogowych dla osób z niepełnosprawnościami stanowią wyzwanie wymagające szczególnej wrażliwości i uwzględnienia specyficznych potrzeb odbiorców. Podejście human in the loop, w którym aktywny udział biorą osoby z niepełnosprawnością, jest kluczowe dla zapewnienia, że opracowywane narzędzie jest dopasowane do ich indywidualnych możliwości. Doktorantka prowadziła testy z udziałem pacjentów, a także wykorzystywała scenariusze odgrywane przez aktorów oraz analizowała wcześniej zebrane dane. Takie podejście pozwala na bieżąco weryfikować skuteczność i użyteczność systemu, a także dostosowywać go do zmieniających się wymagań. O dojrzałość podejścia badaczki świadczy fakt, że celem nie jest zastąpienie specjalisty, lecz wsparcie jego pracy poprzez oszczędność czasu i zwiększenie efektywności interwencji. System ma stanowić narzędzie w rękach terapeuty, a nie jego substytut.

W pracy zdefiniowano cztery cele badawcze. Pierwszy cel badawczy – zbadanie, czy śledzenie wzroku wymaga kalibracji, gdy jest używane w systemie dialogowym – został osiągnięty. Przeprowadzone eksperymenty wykazały brak istotnych różnic w danych dotyczących średnicy źrenicy zebranych z sesji z kalibracją i bez niej. Wyniki te zostały opublikowane w dwóch artykułach naukowych.

Drugi cel badawczy – analiza obszarów zainteresowania w graficznym interfejsie systemu dialogowego – został osiągnięty poprzez zidentyfikowanie i przeanalizowanie obszarów zainteresowania w interfejsie systemu, a następnie rozważenie wykorzystania danych śledzenia wzroku do aktywacji odpowiedzi systemu. Choć początkowa propozycja aktywacji odpowiedzi po wykryciu spojrzenia pacjenta w ustalonym obszarze okazała się problematyczna ze względu na specyfikę zaburzeń psychicznych (częste unikanie kontaktu wzrokowego), analiza ta dostarczyła cennych informacji na temat interakcji pacjenta z systemem.

Trzeci cel badawczy – poprawa płynności rozmów z systemem dialogowym za pomocą śledzenia ruchu gałek ocznych – został osiągnięty dzięki udokumentowanemu pozytywnemu wpływowi tych danych na działanie udoskonalonej wersji systemu Terabot. Przeprowadzone eksperymenty potwierdziły, że wykorzystanie śledzenia wzroku pozwoliło rozwiązać problemy z płynnością rozmów zidentyfikowane podczas wcześniejszych testów. Wyniki te zostały opublikowane w artykule naukowym.

Czwarty cel badawczy – umożliwienie automatycznej oceny zaangażowania użytkownika podczas interakcji z systemem dialogowym – również został osiągnięty. Opracowano rozwiązanie, które tworzy ujednoczoną tablicę zaangażowania pacjenta na podstawie danych z systemu śledzenia wzroku i transkrypcji nagrań audio. W przeciwieństwie do subiektywnych i niejednorodnych raportów sporządzanych przez asystentów, jednolita forma generowana automatycznie znacznie ułatwia dalsze analizy.

Analiza danych offline, zarejestrowanych wcześniej, jak i interakcji z użytkownikami podczas testów wykonywanych na bieżąco, potwierdza słuszność postawionej tezy. Wykazano, że wykorzystanie technologii śledzenia ruchu gałek ocznych poprawia działanie głosowego systemu dialogowego do celów terapeutycznych. Stanowi to istotny krok na przód w rozwoju interaktywnych narzędzi wspierających zdrowie psychiczne.

Uznając istotne zalety przedstawionej rozprawy, pragnę również wskazać kilka zagadnień merytorycznych, które moim zdaniem mogą być przedmiotem konstruktywnej dyskusji.

1. Z uwagi na specyfikę potencjalnych zastosowań, istotne wydaje się doprecyzowanie kwestii dostępu do Internetu przez nowego agenta, opartego na dużym modelu językowym. W przypadku konieczności takiego dostępu, wskazane byłoby przedstawienie mechanizmów zapewniających poufność wrażliwych danych pacjentów. Natomiast w

przypadku funkcjonowania agenta w trybie offline, istotne jest doprecyzowanie kwestii wydajności, czasu reakcji modelu oraz potencjalnych wymagań sprzętowych, w szczególności w odniesieniu do karty graficznej.

2. Charakterystyczny dla dużych modeli językowych brak transparentności algorytmicznej (wytlumaczalności sposobu działania) może rodzić szereg obaw dotyczących potencjalnego, negatywnego wpływu na pacjenta. Czy rozważano ten problem podczas konstrukcji nowej wersji systemu? Jakie mechanizmy zapewniające bezpieczeństwo i dobrostan pacjenta można by zastosować?
3. W podrozdziale 2.2.2, poświęconym systemom dialogowym w obszarze zdrowia psychicznego, omówiono jedynie wybrane rozwiązania bez wskazania kryteriów selekcji. Uzasadnienie motywacji stojącej za tą selekcją pozwoliłoby na pełniejsze zrozumienie kontekstu przeprowadzonych analiz i wyciągniętych wniosków.
4. W odniesieniu do wspomnianego przez autorkę zjawiska nadwrażliwości systemu na zmiany sygnałów sterujących, znanego w literaturze jako efekt Midasa, pojawia się pytanie, czy mimowolna detekcja skupienia uwagi na obszarze twarzy odbiorcy podczas krótkiej pauzy w wypowiedzi pacjenta nie mogłaby prowadzić do przedwczesnego zakończenia wypowiedzi?
5. Czy interfejs użytkownika oraz reprezentacja wizualna agenta w nowej wersji uległy zmianie? Czy rozważano użycie awatara zamiast filmu? Znane są zalety zastosowania nagrania wideo, szczególnie w kontekście zjawiska tzw. doliny niesamowitości, jednak pojawia się pytanie, czy maseczka na twarzy w nagraniu nie wywołuje u użytkownika dyskomfortu?
6. Na rysunku 37 zaprezentowano schemat nowego systemu dialogowego. Warto byłoby przedstawić krótkie uzasadnienie dla wyboru konkretnych komponentów rozpoznawania i syntezy mowy. Czy komponenty te wymagają dostępu do Internetu oraz czy uwzględniono potencjalne implikacje dla bezpieczeństwa danych podawanych przez pacjentów podczas sesji terapeutycznych. Jaki tekst generowany jest przez Fixation signal handling module?
7. Zasygnalizowany przez doktorantkę problem przejmowania inicjatywy przez pacjenta lub przez agenta w trakcie trwającej konwersacji wiąże się z zagadnieniem płynności interakcji, rozwiązywanym z wykorzystaniem śledzenia ruchów gałki ocznej. Czy zapalenie się wskaźnika nagrywania dźwięku, o którym wspomniano w podrozdziale 5.4, potencjalnie zakłócającego pole widzenia, mogłoby być interpretowane jako swego rodzaju wizualny „semafor” sygnalizujący możliwość rozpoczęcia wypowiedzi? Czy takie podejście wydaje się zasadne i uprawnione, w szczególności w kontekście rozważanej grupy pacjentów?

8. W odniesieniu do metodologii zastosowanej w pierwszym zadaniu badawczym, mającym na celu sprawdzenie czy urządzenie do śledzenia ruchu gałek ocznych wymaga kalibracji w głosowym systemie dialogowym, pojawia się pytanie dotyczące stanu początkowego urządzenia. Czy pomiędzy poszczególnymi próbami badawczymi urządzenie było wyłączane? Jeśli tak, czy po ponownym włączeniu zachowywało ustawienia i wyniki poprzedniej kalibracji?

Mam także kilka drobnych uwag redakcyjnych.

1. Strona 28: „... AI-based CAs are described from other countries.” - być może zgrabniej byłoby „... AI-based CAs from other countries are described.”
2. Strona 35: Pojawia się pojedyncza pozycja listy wypunktowanej „Eye-tracking devices – design”.
3. Strona 39: Wspomina się o sakadach, podczas gdy wyjaśnienie tego zjawiska jest później.
4. Strona 52: W opisie rysunku występuje drobna nieścisłość. Rysunek nie przedstawia pacjenta ani interfejsu Terabot na ekranie, mimo że w opisie napisano: „...as shown in the figure”.

Przedstawione w punktach 1-8 uwagi wynikają z chęci pogłębienia dyskusji nad prezentowanymi zagadnieniami i nie stanowią zarzutów. Nieliczne usterki nie wpływają na pozytywny odbiór i nie umniejszają oryginalności i innowacyjności udokumentowanych w rozprawie osiągnięć.

4 Wniosek końcowy

Rozprawa doktorska prezentuje ogólną wiedzę teoretyczną doktorantki w dyscyplinie informatyka techniczna i telekomunikacja oraz stanowi potwierdzenie umiejętności samodzielnego prowadzenia pracy naukowej. Jej przedmiotem jest oryginalne rozwiązanie, wykorzystujące wyniki własnych badań naukowych w sferze społecznej.

Biorąc pod uwagę powyższe, stwierdzam, że przedstawiona do recenzji rozprawa doktorska autorstwa Pani mgr inż. Karoliny Teresy Gabor-Siatkowskiej pt. Improving Therapeutic Spoken Dialogue Systems with Eye Tracking spełnia wymogi formalne i wnoszę o jej przyjęcie i dopuszczenie do dalszych etapów postępowania doktorskiego.

Tomasz Kupuskiński



Politechnika
Wroclawska

Wroclaw, 8.12.2025 r.

**Recenzja pracy doktorskiej mgr inż. Karoliny Gabor-Siatkowskiej zatytułowanej:
„Improving Therapeutic Spoken Dialogue Systems with Eye Tracking (Ulepszanie
terapeutycznych głosowych systemów dialogowych przy użyciu okulografii)”
przygotowanej pod kierunkiem: dra hab. inż. Artura Janickiego, prof. uczelni w
Politechnice Warszawskiej; w dyscyplinie Informatyka Techniczna i Telekomunikacja.**

Recenzja została sporządzona w związku z powołaniem przez przewodniczącego Rady Naukowej Dyscypliny Informatyka Techniczna i Telekomunikacji pana prof. dra hab. inż. Jarosława Arabasa na posiedzeniu w dniu 23.09.2025 roku do pełnienia funkcji recenzentki w postępowaniu o nadanie stopnia naukowego doktora nauk technicznych w dyscyplinie Informatyka Techniczna i Telekomunikacja pani mgr inż. Karolinie Gabor-Siatkowskiej.

Niniejsza recenzja ma za zadanie zgodnie z art. 186 ust. 1 ustawy z dnia 20 lipca 2018 r. – Prawo o szkolnictwie wyższym i nauce, w celu oceny, czy rozprawa doktorska stanowi oryginalne rozwiązanie problemu naukowego i wnosi istotny wkład w rozwój dyscypliny naukowej Informatyka Techniczna i Telekomunikacja.

W ramach przeprowadzonej recenzji zostaną ocenione następujące punkty:

1. Tematyka pracy doktorskiej i jej wkład w dyscyplinę Informatyka Techniczna i Telekomunikacja.

Ad. 1. Temat pracy doktorskiej brzmi: „Improving Therapeutic Spoken Dialogue Systems with Eye Tracking”.

W ramach realizacji niniejszej rozprawy doktorskiej opracowano system dialogu głosowego oparty na LLM (Terabot) z wykorzystaniem ruchu gałek ocznych, przeznaczony do pomocniczej terapii w przebiegu schizofrenii. Opracowany system opiera się na badaniach pilotażowych przeprowadzonych w IPIN Warsaw, w których zidentyfikowano problemy z płynnością dialogu za pomocą danych z eyetrackingu pacjentów. Przeprowadzona przez Doktorantkę analiza eyetrackingu ujawniła obszary zainteresowania pacjentów podczas mówienia, umożliwiając multimodalną aktualizację z wykorzystaniem modeli GPT, które dostosowują rozmowy do ruchu gałek ocznych, poprawiając naturalność i radząc sobie z pauzami w mowie podczas oceny zaangażowania. Opracowany system wspiera cyfrowo zdrowie psychiczne poprzez integrację sygnałów

niewerbalnych z LLM, przewyższając systemy oparte wyłącznie na mowie w realistycznej interakcji z pacjentem.

W mojej ocenie – wybrany temat rozprawy o charakterze wybitnie interdyscyplinarnym, jest aktualny i stanowi istotny wkład dla dyscypliny naukowej: **Informatyka Techniczna i Telekomunikacja**.

2. Zagadnienia naukowe rozprawy – teza.

W pracy została postawiona następująca teza główna badawcza:

- „*The use of eye tracking can improve a therapeutic spoken dialogue system (Wykorzystanie śledzenia ruchu gałek ocznych może poprawić terapeutyczny system dialogu mówionego)*”.

Oraz **4** cele badawcze (Research Objectives - RO):

- **RO1:** Investigating whether an eye tracker needs to be calibrated when used in a spoken dialogue system (Badanie, czy śledzenie wzroku wymaga kalibracji w przypadku korzystania z systemu dialogów mówionych).
- **RO2:** Analysis of areas of interest in the graphical interface of the dialogue system (Analiza obszarów zainteresowania w interfejsie graficznym systemu dialogowego).
- **RO3:** Enhancing the fluency of conversations with a dialogue system by the use of an eye tracker (Poprawa płynności rozmów dzięki systemowi dialogowemu wykorzystującemu śledzenie ruchu gałek ocznych).
- **RO4:** Enabling automatic user engagement assessment while interacting with the dialogue system (Włączanie automatycznej oceny zaangażowania użytkownika podczas interakcji z systemem dialogowym).

Biorąc pod uwagę osiągnięte przez Doktorantkę wyniki - postawiona w pracy teza oraz cele badawcze zostały udowodnione.

3. Struktura pracy

Ad. 3. Niniejsza rozprawa doktorska została w całości napisana w języku angielskim i składa się ze **120** stron (wliczając stronę tytułową, podziękowania, streszczenia - w języku polskim oraz angielskim, spis treści, bibliografię, załączniki, spis rysunków oraz tabel) i została podzielona na **8** ponumerowanych rozdziałów, załączniki (**A** i **B**), spis rysunków oraz tabel, nieponumerowaną i niewymienioną w spisie treści bibliografię. W pracy zawarto **11** tabel (wliczając **2** tabele w załącznikach) oraz **46** rysunków. Załączniki zawierają przykład dialogu - wybrany temat: złość (Załącznik **A**) oraz scenariusze dla aktorek i aktorów "Teatru PW" (Załącznik **B**). Bibliografia składa się ze **138** pozycji, w skład w których wchodzi **5** publikacji Autorki rozprawy.

Rozdział pierwszy zatytułowany "Introduction (Wprowadzenie)" zawiera ogólne wprowadzenie do tematu, informacje na temat dorobku Doktorantki oraz jej osiągnięć

naukowych. **Rozdział drugi**, który stanowi przegląd literatury, omawia wyzwania i znaczenie opieki psychiatrycznej na świecie. Doktorantka skupia się w nim także na agentach konwersacyjnych, wyjaśniając główne komponenty systemów dialogowych i ich zastosowanie w cyfrowej opiece zdrowotnej, ze szczególnym uwzględnieniem rozwiązań opartych na dużych modelach językowych (LLM) wspierających dobrostan psychiczny. W rozdziale tym przedstawiono także technologię śledzenia wzroku, omawiając parametry techniczne urządzeń i ich wykorzystanie w interakcji człowiek-komputer. W **rozdziale trzecim** opisuje stworzony przez Autorkę system Terabot dla pacjentów z zaburzeniami psychicznymi w języku polskim, wraz z przeprowadzonymi eksperymentami klinicznymi w Instytucie Psychiatrii i Neurologii. W rozdziale tym omówiono także architekturę systemu, interfejs użytkownika oraz eye-tracker Gazepoint GP3 użyty podczas badań. Zidentyfikowano kluczowe problemy dialogu oraz oceniono satysfakcję pacjentów z rozmów z Terabotem. **Rozdział czwarty** analizuje potrzebę integracji eye-trackera z systemem dialogowym, opisując problemy dialogu zidentyfikowane w Terabocie. Przedstawiono w nim także metodologię eksperymentu, wyniki badań oraz dyskusję, kończąc wnioskami na temat konieczności wykorzystania danych wzrokowych do poprawy naturalności rozmów z pacjentami. W **rozdziale piątym** Doktorantka analizuje wzorce wzroku pacjentów podczas rozmów z Terabotem, opisując problem, metodologię badań oraz analizę danych z eye-trackera zsynchronizowanych z wypowiedziami. Przedstawiono w nim także otrzymane wyniki badań oraz dyskusję na temat klinicznego zastosowania proponowanego rozwiązania w terapii zaburzeń psychicznych. **Rozdział szósty** zawiera opisy dotyczące projektowania systemu dialogowego zależnego od wzroku, opartego na LLM (OpenAI Assistant API), dla celów terapeutycznych. Omówiono analizę czasów oczekiwania pacjentów, wykorzystanie danych eye-tracking w czasie rzeczywistym, testy offline na danych pacjentów oraz testy real-time z symulowanymi pacjentami i aktorami Teatru PW. Przedstawiono w nim wyniki analizy wydajności nowego systemu, w tym aktywację reakcji na eye-tracking, ocenę rozmów oraz ankietę satysfakcji uczestników. **Rozdział siódmy** przedstawia automatyczną ocenę zaangażowania pacjenta w dialogi terapeutyczne. Omówiono analizę notatek asystentów klinicznych dotyczących zaangażowania pacjentów, zaproponowano metodę automatycznej oceny wykorzystującą dane eye-tracking oraz przedstawiono wyniki badań i wnioski dotyczące tego rozdziału. **Rozdział ósmy** zawiera podsumowanie i dyskusję wyników badań. Autorka omówiła w nim ograniczenia pracy oraz kierunki dalszych badań. Ponadto praca zawiera załączniki A i B, spis tabel, spis rysunków oraz bibliografię składającą się ze **138** pozycji.

4. Uwagi redakcyjne, krytyczne oraz pytania do pracy.

Ad. 4.

- Uwagi krytyczne:
 - Str. 1.:
 - Z rozdziału do podrozdziału - bez wprowadzenia.
 - Rysunek 1 - lekko rozmazana czcionka, słaba jakość.
 - Str. 2.:
 - Rysunek 2 - j.w.
 - Str. 11.:

- Z rozdziału do podrozdziału - bez wprowadzenia.
- Str. 12.:
 - Rys. 3. - j.w.
- Str. 18.:
 - Rys. 5. - j.w.
- Str. 20.:
 - Rys. 6 - j.w.
- Str. 27.:
 - Rys. 14. - j.w.
- Str. 34.:
 - Z rozdziału do podrozdziału - bez wprowadzenia.
- Str. 46.:
 - Rys. 38. - j.w.
- Str. 47.:
 - Z rozdziału do podrozdziału - bez wprowadzenia.
- Str. 48.:
 - Opis eksperymentu badawczego byłby bardziej czytelny w formie tabeli oraz/lub wykresu.
- Str. 54.:
 - Z rozdziału do podrozdziału - bez wprowadzenia.
- Str. 61.:
 - Z rozdziału do podrozdziału - bez wprowadzenia.
- Mała próba badawcza: Analizy eye-tracking na danych zaledwie 9 pacjentów ze schizofrenią, co negatywnie wpływa na moc statystyczną.
- Brak porównań ilościowych w pracy, gdzie nie pokazano przewagi multimodalnego Terabot nad innymi systemami dialogowymi (np. ELIZA, Woebot) w metrykach jak BLEU, ROUGE czy engagement score. Uzyskane wyniki offline nie znajdują statystycznego potwierdzenia w rozprawie.
- Pozytywne aspekty pracy:
 - Rozprawa wyróżnia się pięknym, poprawnym językiem angielskim oraz wyjątkową dbałością o szczegóły edytorskie.
 - Cała praca jest niezwykle precyzyjnie dopracowana, a nawet na rysunkach z dialogami w języku polskim konsekwentnie podano wyraźne tłumaczenia na język angielski, co świadczy o profesjonalizmie i spójności całej pracy.
 - Autorka opracowała innowacyjne połączenie eye-tracking z LLM (GPT-4o-mini + RAG) w terapii schizofrenii, gdzie zgodnie z moim stanem wiedzy, jest to pierwszy system reagujący na spojrzenie pacjenta w czasie rzeczywistym.
 - Autorski system Terabot był testowany klinicznie i wdrażany testowo w Instytucie Psychiatrii i Neurologii (IPiN) Warszawa, co znacząco wzmacnia wiarygodność proponowanych rozwiązań.
- Pytania do pracy:
 - Jakie są, Pani zdaniem, główne ograniczenia przeprowadzonych badań i, które z nich uważa Pani za najbardziej krytyczne?

- Jakie mechanizmy kontroli bezpieczeństwa wypowiedzi systemu opartego na LLM zostały zastosowane lub byłyby konieczne w warunkach rzeczywistej terapii?
- Czy uzyskane przez Panią wyniki na grupie pacjentów ze schizofrenią można przenieść na inne grupy kliniczne dotknięte zaburzeniami psychicznymi?

5. Podsumowanie.

Ad. 5. Niniejsza rozprawa została niezwykle starannie zredagowana i estetycznie opracowana, bez istotnych błędów redakcyjnych. Ogólna dbałość o detale znacząco podnosi wartość pracy. Recenzowanie tej pracy było prawdziwą przyjemnością, gdyż napisana ona była bardzo dobrym językiem angielskim, starannie, a tematyka jest niezwykle ciekawa i aktualna.

Krytyczne uwagi mają charakter głównie kosmetyczny, natomiast zadane pytania mają charakter pogłębiający i dociekliwy. Z trudem byłam w stanie znaleźć niedociągnięcia czy negatywy w tej pracy.

Na uwagę zasługuje także całokształt osiągnięć i działań naukowych Doktorantki. Cenne jest szczegółowe opisanie przez Autorkę swoich publikacji, osiągnięć oraz udziału w projektach krajowych i międzynarodowych. Rzadko spotyka się taki opis w pracach doktorskich, a przecież to właśnie taki pełny autoreferat pozwala na kompleksową ocenę wkładu w dyscyplinę i świadczy o aktywnym udziale w życiu naukowym Doktorantki.

Autorka jest autorką/współautką **13** publikacji naukowych (Google Scholar) i posiada następujący indeks Hirscha: **4** (Google Scholar), **47** cytowań. W **10** z **13** publikacji Doktorantka jest pierwszą (wiodącą) autorką.

Doktorantka opracowała rozwiązania problemów dialogu (przerywanie, pauzy) wykorzystując dane wzrokowe z GPT-4o-mini + RAG, potwierdzone testami real-time i automatyczną oceną zaangażowania. Posiada w swoim dorobku aż **13** publikacji naukowych, brała udział w projektach, otrzymywała nagrody Rektora, co świadczy o wpływie na rozwój dyscypliny jaką jest **Informatyka Techniczna i Telekomunikacja**.

W mojej ocenie rozprawa doktorska **mgr inż. Karoliny Gabor-Siatkowskiej** zawiera bardzo wartościowe wyniki badań, wyróżniając się potencjalną wartością praktyczną - potencjałem na wdrożenie klinicznie opracowanego systemem Terabot; a moja ocena jest jak najbardziej **pozytywna** - praca prezentuje cenne osiągnięcia naukowe w dyscyplinie **Informatyka Techniczna i Telekomunikacja**, spełniając wszystkie wymogi art. 186 ust. 1 ustawy z 20 lipca 2018 r. - Prawo o szkolnictwie wyższym i nauce, i nadaje się do publicznej obrony.

Ze względu na osiągnięte i zaprezentowane w pracy wyniki, a także na bardzo dobry dorobek naukowy Doktorantki, jej bardzo dużą (na tym etapie kariery naukowej) aktywność, wnioskuję o **wyróżnienie rozprawy mgr inż. Karoliny Gabor-Siatkowskiej**.

Aleksandra Kawala Sterniuk

Dr hab. inż. Aleksandra Kawala-Sterniuk, prof. uczelni
Katedra Sztucznej Inteligencji
Wydział Informatyki i Telekomunikacji
Politechnika Wrocławska
aleksandra.kawala-sterniuk@pwr.edu.pl