

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 musiałby 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

