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 umożliwia 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 ujednoliconą 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 naprzód 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 (wytłumaczalności 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 zapalanie 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.