

RECENZJA ROZPRAWY DOKTORSKIEJ

Autor rozprawy: mgr Dawid Ratajczyk

Tytuł rozprawy: *Perceiving Artificial Agents: A Multimethod Investigation into the Shape and Variables of the Uncanny Valley*

Jednostka: Wydział Psychologii i Kognitywistyki, Uniwersytet im. Adama Mickiewicza w Poznaniu

Promotor: dr hab. Paweł Łupkowski, prof. UAM

Promotor pomocniczy: dr inż. Marcin Jukiewicz

1. Wprowadzenie

Przedstawiona do recenzji rozprawa doktorska dotyczy jednego z najbardziej dyskutowanych zagadnień współczesnych badań nad interakcją człowieka z technologią, a mianowicie zjawiska określanego w literaturze jako „dolina niesamowitości” (uncanny valley). Termin ten odnosi się do specyficznej reakcji emocjonalnej pojawiającej się w sytuacji kontaktu z artefaktami technologicznymi – takimi jak roboty lub wirtualni agenci – które wykazują wysoki, lecz niepełny stopień podobieństwa do ludzi. W takich sytuacjach obserwatorzy często doświadczają uczucia niepokoju, dziwności lub dyskomfortu, które nie pojawia się ani w przypadku wyraźnie nieludzkich obiektów, ani w kontakcie z rzeczywistymi ludźmi.

Problem badawczy podjęty w rozprawie jest bez wątpienia istotny zarówno z punktu widzenia badań nad robotyką społeczną i interakcją człowiek–komputer, jak i z perspektywy ogólniejszych badań nad poznaniem społecznym. Jednocześnie należy zauważyć, że literatura dotycząca uncanny valley jest w znacznym stopniu rozproszona, a proponowane wyjaśnienia teoretyczne często koncentrują się na wąskich aspektach zjawiska. W wielu pracach badawczych nacisk kładzie się przede wszystkim na kwestie związane z percepcyjnym niedopasowaniem cech wizualnych lub ruchowych, natomiast rzadziej uwzględnia się szerszy kontekst mechanizmów poznania społecznego.

Na tym tle rozprawa mgr. Dawida Ratajczyka stanowi interesującą próbę poszerzenia perspektywy badawczej. Autor wskazuje na znaczenie takich czynników jak przekonania esencjalistyczne dotyczące natury ludzkiej oraz interakcja pomiędzy wyglądem sztucznego agenta a przypisywaniem mu stanów mentalnych. Dzięki temu badania przedstawione w rozprawie wpisują problem doliny niesamowitości w szerszy kontekst badań nad poznaniem społecznym. Uważam, że jest to jedno z najważniejszych osiągnięć całego projektu badawczego.

2. Charakterystyka ogólna rozprawy

Rozprawa ma formę cyklu artykułów naukowych poprzedzonych obszernym wprowadzeniem teoretycznym. Taka forma pracy doktorskiej jest obecnie szeroko akceptowana w naukach empirycznych i umożliwia prezentację wyników badań w postaci publikacji w międzynarodowych czasopismach naukowych. W przypadku omawianej rozprawy dwa artykuły zostały już opublikowane, natomiast trzeci został przyjęty do publikacji. Biorąc pod uwagę rangę czasopism oraz długość procesu wydawniczego, należy uznać to za znaczące osiągnięcie doktoranta.

Wprowadzenie do rozprawy zostało przygotowane bardzo starannie. Autor w klarowny sposób przedstawia logikę całego programu badawczego oraz pokazuje, w jaki sposób kolejne artykuły rozwijają wcześniejsze ustalenia. Dzięki temu czytelnik otrzymuje spójny obraz projektu badawczego, w którym poszczególne badania pełnią komplementarne role.

Na szczególne podkreślenie zasługuje również analiza problemów związanych z operacjonalizacją zmiennych w badaniach nad uncanny valley. Autor pokazuje, że takie pojęcia jak eeriness, likability czy familiarity bywają stosowane w literaturze w sposób niejednoznaczny. Próba uporządkowania tej terminologii stanowi wartościowy wkład w dyskusję metodologiczną dotyczącą badań nad reakcjami ludzi na sztucznych agentów.

3. Ocena pierwszego artykułu

Pierwszy artykuł wykorzystuje analizę dużego zbioru komentarzy pochodzących z serwisu YouTube. Podejście to jest interesujące metodologicznie, ponieważ pozwala analizować spontaniczne reakcje użytkowników internetu na różne typy robotów. Analiza materiałów powstających w naturalnych kontekstach komunikacyjnych może stanowić cenne uzupełnienie badań laboratoryjnych, które często opierają się na ograniczonych próbach badawczych.

Jednocześnie należy zauważyć, że wyniki dotyczące kształtu zależności pomiędzy humanlikeness a reakcjami emocjonalnymi nie są w pełni przekonujące. W szczególności wartości kryterium AIC_c dla modeli liniowych, kwadratowych oraz sześciennych są w wielu przypadkach bardzo zbliżone. W efekcie trudno jednoznacznie rozstrzygnąć, który model najlepiej opisuje analizowane zależności. Jedynie w przypadku analiz dotyczących cech twarzy można mówić o wyraźniejszym wsparciu dla modelu nieliniowego.

Problem ten wydaje się być w dużej mierze związany z ograniczoną liczbą robotów uwzględnionych w analizie. Ograniczenie to wynikało z chęci zapewnienia odpowiednio dużej liczby komentarzy dla każdego robota, co jest zrozumiałe z punktu widzenia rzetelności analizy językowej. Jednocześnie jednak prowadzi ono do sytuacji, w której

badanie dotyczy reakcji na stosunkowo niewielką liczbę robotów, co powoduje, że stosunkowo duża wariancja pomiędzy poszczególnymi przypadkami znacznie wpływa na analizę.

W konsekwencji wnioski dotyczące dokładnego kształtu krzywej uncanny valley wydają się słabiej uzasadnione empirycznie niż sugeruje to dyskusja teoretyczna. Dotyczy to między innymi wniosku, że zjawisko to nie jest powiązane z mechanizmami doboru seksualnego. Brak związku pomiędzy atrakcyjnością a humanopodobieństwem nie musi bowiem oznaczać, że mechanizmy związane z doбором seksualnym nie odgrywają żadnej roli. Aby sformułować tak daleko idący wniosek należałoby między innymi wykazać, że reakcje typu eeriness nie pełnią funkcji sygnału ostrzegawczego wskazującego, że obiekt wywołujący zainteresowanie nie jest wartościowym partnerem reprodukcyjnym. Interesującym kierunkiem dalszych badań mogłoby być również sprawdzenie, czy struktura reakcji uncanny valley różni się pomiędzy kobietami i mężczyznami, co można byłoby powiązać z różnymi kosztami inwestycji reprodukcyjnej.

Kolejną kwestią metodologiczną jest wykorzystanie bazy danych ABOT. Oceny humanopodobieństwa w tej bazie opierają się na statycznych obrazach robotów. Tymczasem w analizowanych materiałach wideo istotne znaczenie może mieć sposób poruszania się robotów. W szczególności skala body-manipulators odnosi się jedynie do obecności kończyn i elementów manipulacyjnych, natomiast nie uwzględnia sposobu, w jaki roboty wykonują ruchy przypominające ludzkie zachowania. Jest zatem bardzo prawdopodobne, że część reakcji użytkowników analizowanych w komentarzach wynika z aspektów zachowania robotów, które nie są uchwycone przez zastosowane miary.

Na marginesie warto również wspomnieć o pewnej niejasności dotyczącej liczby robotów analizowanych w badaniu. W tekście artykułu pojawia się informacja o trzydziestu trzech robotach, natomiast lista w jednym z aneksów zawiera jedynie dwadzieścia siedem pozycji. Różnica ta nie została wyjaśniona w tekście i mogłaby zostać doprecyzowana.

Pomimo wskazanych ograniczeń artykuł stanowi interesującą próbę wykorzystania danych pochodzących z naturalnych kontekstów komunikacyjnych do badania zjawiska uncanny valley. Autor podejmuje ambitny wysiłek wydobycia znaczących wzorców z materiału empirycznego, który sam w sobie jest bardzo zróżnicowany i trudny do interpretacji.

4. Ocena drugiego artykułu

Drugi artykuł przedstawia dobrze zaplanowane badanie eksperymentalne analizujące wpływ dwóch czynników: sposobu prezentacji bodźców (wirtualna rzeczywistość versus ekran komputera) oraz przekonań dotyczących wyjątkowości natury ludzkiej. Jednym z ważniejszych rezultatów jest brak istotnych różnic pomiędzy warunkami VR i prezentacją na ekranie komputera. Wynik ten ma znaczenie metodologiczne, ponieważ sugeruje, że w wielu badaniach nad uncanny valley można stosować prostsze paradygmaty eksperymentalne bez istotnej utraty trafności wyników.

Warto jednak zauważyć, że wszystkie postacie wykorzystane w badaniu wykazywały identyczne wzorce ruchu i zachowania. Było to zrozumiałe z punktu widzenia kontroli eksperymentalnej, jednak ruch sam w sobie stanowi istotną zmienną w badaniach nad uncanny valley. Inne zestawy zachowań mogłyby wchodzić w interakcje z wyglądem (analogicznie do badania w artykule numer 3) i potencjalnie prowadzić do odmiennych ocen niesamowitości. Jest to ważny aspekt, który mógłby zostać szerzej omówiony w kontekście interpretacji wyników.

Inną kwestią jest fakt, że badanie porównuje reakcje w środowisku VR i na ekranie komputera, nie obejmuje natomiast interakcji z fizycznymi robotami. Choć jest oczywiste, dlaczego nie było to częścią metodologii tego badania, warto byłoby szerzej omówić tę kwestię w kontekście ograniczeń wyników.

Najważniejszym rezultatem artykułu jest jednak wykazanie związku pomiędzy reakcjami niesamowitości a przekonaniem esencjalistycznymi dotyczącymi natury ludzkiej. Wynik ten wskazuje, że uncanny valley może być powiązana z bardziej ogólnymi mechanizmami poznania społecznego. Otwiera to interesującą perspektywę interpretacyjną, która wykracza poza dominujące w literaturze wyjaśnienia oparte wyłącznie na percepcyjnym niedopasowaniu bodźców.

W tym kontekście warto zauważyć, że literatura dotycząca poznania społecznego oferuje szeroki zestaw potencjalnych mechanizmów, które mogłyby zostać wykorzystane do interpretacji zjawiska uncanny valley. Należą do nich między innymi procesy mentalizacji, naruszanie oczekiwań społecznych, mechanizmy predykcyjnego przetwarzania informacji, zaburzenia wspólnej uwagi czy regulacja granic kategorii społecznych. Androidy i wirtualni agenci mogą stanowić szczególnie interesujące narzędzia eksperymentalne do badania tych mechanizmów, ponieważ umożliwiają systematyczne manipulowanie stopniem podobieństwa do ludzi przy jednoczesnej kontroli wielu innych czynników.

W tym sensie uncanny valley może być traktowana nie tylko jako problem dotyczący reakcji ludzi na roboty, lecz także jako narzędzie do badania architektury poznania społecznego w sytuacjach, w których granice kategorii „człowiek” i „nie-człowiek” ulegają częściowemu rozmyciu.

5. Ocena trzeciego artykułu

Trzeci artykuł stanowi najbardziej rozwinięty element całej rozprawy i wyraźnie pokazuje rozwój kompetencji metodologicznych autora. Badanie analizuje interakcję pomiędzy wyglądem agenta a przypisywanym mu typem umysłu. Wykorzystano zarówno miary deklaratywne, jak i pomiary fizjologiczne, co pozwala uzyskać bardziej wielowymiarowy obraz reakcji uczestników.

Zastosowanie pomiarów fizjologicznych należy uznać za bardzo wartościowy element metodologii. Szczególnie interesujące jest to, że różne wskaźniki wykazały odmienne wzorce wyników. Tego rodzaju rozbieżności pomiędzy miarami deklaratywnymi a fizjologicznymi są często obserwowane w badaniach psychologicznych i mogą dostarczać cennych informacji na temat złożoności analizowanych procesów.

W badaniu zastosowano wcześniej nagrane odpowiedzi odtwarzane uczestnikom w trakcie interakcji. Rozwiązanie to było zrozumiałe z punktu widzenia potrzeby zachowania kontroli eksperymentalnej oraz uniknięcia różnic pomiędzy warunkami eksperymentalnymi. Jednocześnie można zastanawiać się, czy tego rodzaju interakcja w pełni oddaje sposób, w jaki ludzie reagują na rzeczywistego rozmówcę. Potencjalnie wartościowym rozwiązaniem metodologicznym mogłoby być przeprowadzenie pilotażu porównującego interakcję z nagranyymi odpowiedziami i z rzeczywistym rozmówcą, który stara się odpowiadać w możliwie neutralny sposób.

Pewne pytania budzi również dodatkowy tekst przedstawiany uczestnikom w warunkach związanych z AI, w którym podkreślano zdolność sztucznej inteligencji do empatii. Tego rodzaju informacja może wzmacniać procesy mentalizacji u uczestników i w pewnym stopniu wpływać na ich reakcje. Nie jest do końca jasne, dlaczego uznano tę manipulację za konieczną.

Inną kwestią jest sposób operacjonalizacji humanopodobieństwa poprzez pytanie o stopień podobieństwa modelu do człowieka. Ponieważ uczestnicy byli wcześniej informowani o charakterze interakcji, mogli interpretować to pytanie w różny sposób w zależności od warunków eksperymentalnych. Na przykład w sytuacji, w której robotyczny awatar miał być sterowany przez człowieka, uczestnicy mogli interpretować pytanie jako odnoszące się jedynie do wyglądu postaci. Rezultaty badań jednak nie wskazują na taki efekt.

Warto również zauważyć, że postać w warunku „human appearance” była również generowana komputerowo, a więc potencjalnie mogła sama wywoływać efekt uncanny valley. Wyniki artykułu numer 2 sugerują jednak, że tego rodzaju postacie wywołują znacznie słabsze reakcje, dzięki czemu mogą stanowić rozsądny punkt odniesienia dla porównań.

6. Ocena całościowa

Cykl artykułów składających się na rozprawę pokazuje wyraźny rozwój zarówno pod względem teoretycznym, jak i metodologicznym. Kolejne badania stopniowo rozwijają wcześniejsze pomysły i prowadzą do coraz bardziej precyzyjnych pytań badawczych. Szczególnie istotny jest wkład autora w powiązanie badań nad uncanny valley z szerszym kontekstem badań nad poznaniem społecznym.

Przedstawione badania sugerują, że sztuczni agenci mogą być traktowani nie tylko jako obiekty budzące określone reakcje emocjonalne, lecz także jako narzędzia eksperymentalne umożliwiające badanie podstawowych mechanizmów funkcjonowania ludzkiego umysłu w sytuacjach społecznych. W tym sensie praca wpisuje się w rosnący nurt badań wykorzystujących roboty i awatary jako narzędzia badawcze w naukach kognitywnych.

Cykl artykułów pokazuje również wyraźny rozwój kompetencji metodologicznych autora. Pierwszy artykuł ma charakter bardziej eksploracyjny i opiera się na analizie dużych zbiorów danych, natomiast kolejne badania wykorzystują coraz bardziej kontrolowane paradygmaty eksperymentalne. Taka ewolucja programu badawczego świadczy o rosnącej dojrzałości metodologicznej doktoranta.

7. Wniosek końcowy

Biorąc pod uwagę przedstawione powyżej argumenty stwierdzam, że rozprawa doktorska mgr. Dawida Ratajczyka stanowi wartościowy wkład w badania nad zjawiskiem tzw. doliny niesamowitości oraz nad poznawczymi podstawami interakcji człowieka ze sztucznymi agentami. Przedstawiony cykl artykułów pokazuje wyraźny rozwój zarówno pod względem teoretycznym, jak i metodologicznym, a także zdolność autora do prowadzenia badań naukowych oraz publikowania ich wyników w międzynarodowych czasopismach naukowych.

Należy przy tym zaznaczyć, że jeden z artykułów wchodzących w skład rozprawy nie został jeszcze opublikowany, lecz został przyjęty do druku w czasopiśmie naukowym. W mojej ocenie fakt ten nie stanowi merytorycznej przeszkody dla pozytywnej oceny rozprawy. Należy pamiętać, że proces publikacyjny w czasopismach międzynarodowych jest często długotrwały, a przyjęcie artykułu do publikacji stanowi potwierdzenie jego pozytywnej oceny w procedurze recenzyjnej.

W związku z powyższym, o ile obowiązujące w danej jednostce przepisy dopuszczają włączenie do rozprawy artykułu przyjętego do publikacji, uznaję przedstawiony cykl prac za spełniający wymagania stawiane rozprawom doktorskim.

W związku z powyższym stwierdzam, że rozprawa doktorska mgr. Dawida Ratajczyka spełnia wymagania określone w obowiązujących przepisach i wnoszę o dopuszczenie doktoranta do dalszych etapów postępowania doktorskiego.

Białystok, dnia 9/03/2026

A handwritten signature in blue ink, appearing to read 'Konrad Talmont-Kaminski', written in a cursive style.

dr hab. Konrad Talmont-Kaminski, Prof. UwB

Uniwersytet w Białymstoku

W związku z powyższym stwierdzam, że rozprawa doktorska mgr. Dawida Ratajczyka spełnia wymagania określone w obowiązujących przepisach i wnoszę o dopuszczenie doktoranta do dalszych etapów postępowania doktorskiego.

Białystok, dnia 9/03/2026

A handwritten signature in blue ink, appearing to read 'Konrad Talmont-Kaminski', written in a cursive style.

dr hab. Konrad Talmont-Kaminski, Prof. UwB

Uniwersytet w Białymstoku