płk prof. dr hab. inż. Bogdan GRENDA Warszawa, 27.01.2025r.

RECENZJA
rozprawy doktorskiej
pt. „Zagrożenie dezinformacją multimedialną z wykorzystaniem technologii deepfake z perspektywy bezpieczeństwa narodowego" opracowanej przez Pana mgr Bartosza BIDERMANA
pod kierunkiem naukowym Pana prof. dr hab. inż. Tadeusza SZCZURKA oraz promotora pomocniczego Pana dr Jakuba ADAMKIEWICZA

Oświadczenie: Recenzent oświadcza, że nie posiada żadnych interesów w odniesieniu do recenzowanej rozprawy doktorskiej, a tematyka tej dysertacji mieści się w nurcie jego zainteresowań naukowo-badawczych.

Układ rozprawy

Dysertacja zawiera wszystkie wymagane elementy redakcyjne prac naukowych. W skład dysertacji wchodzi: streszczenie w j. polskim i angielskim oraz słowa kluczowe (4 strony), spis treści (4 strony), wprowadzenie (10 stron), sześć rozdziałów merytorycznych (ogółem 315 stron), zakończenie (7 stron), bibliografia (14 stron), spis ilustracji, tabel i wykresów (12 stron), oraz załączniki (38 stron). Praca doktorska to obszerne dzieło liczące prawie 400 stronicowe.

Za poprawny uważam układ recenzowanej pracy doktorskiej. Jej struktura jest przejrzysta, a prowadzone wywody stanowią powiązaną całość. Pracę cechuje pragmatyzm naukowy, bowiem treści poszczególnych jej części są ze sobą powiązane - następne wynikają z poprzednich jako efekt nakreślonego sposobu przedstawiania poszczególnych kwestii i prezentacji uzyskanych wyników badań. Rozprawa zapoczątkowana jest wprowadzeniem, a następnie rozwinięta w sześciu rozdziałach, stanowi zamkniętą i tematycznie spójną całość.

Dysertację otwiera *„Wprowadzenie”*, które stanowi syntetyczny materiał wprowadzający w problematykę opracowania oraz zawiera właściwe streszczenie poszczególnych rozdziałów.

Koncepcja metodologiczna rozprawy została umieszczona w rozdziale pierwszym rozprawy zatytułowanym *„Konceptualizacja badań nad manipulacją obrazem”* (s. 21-55), w którym przedstawiono podstawowe elementy naukowej metodologii badawczej, obejmujące: założenia metodologiczne badań (w tym: sytuację problemową, przedmiot i cel badań, problemy i zmienne w procesie badawczym, hipotezy badawcze), a także prezentację zastosowanych metod badawczych oraz ograniczenia badawcze. Przyjętą koncepcję metodologiczną należy uznać za poprawną, adekwatną do potrzeb rozprawy.

W rozdziale drugim pt.: „Wpływ zmanipulowanych materiałów audiowizualnych na bezpieczeństwo narodowe” – zawarto ustalenia terminologiczne takich pojęć jak: zagrożenie, manipulacja, dezinformacja multimedialna, bezpieczeństwo, bezpieczeństwo narodowe. W dalszej kolejności Autor przedstawił etymologię i ewolucję terminu „deepfake”. Kolejna część poświęcona została ustaleniu zakresu informacji zawartych w przepisach prawnych, a dotyczących filmów deepfake. Doktorant porusza także kwestię roli i znaczenie materiałów wizualnych, a także omawia modele wykorzystania technologii deepfake w kreowaniu wirtualnej rzeczywistości m.in. takich jak: edycja zdjęcia z wykorzystaniem innego zdjęcia, trening audio, zmiana głosu, zmiana twarzy (np. morfing). Wskazuje także na pojęcie „Puppetry deepfake” oznaczające tworzenie realnie wyglądających animacji twarzy lub całego ciała, na podstawie jednego zdjęcia. Autor trafnie zauważa, że *„Zwieńczeniem wszystkich technologii kryjących się lub powiązanych z pojęciem deepfake, jest ich połączenie w spójną całość. Dopiero prawidłowe zgranie audio, jak i wideo daje oglądającemu pełnię przekonania, iż ogląda prawdziwe nagranie. Prawidłowa synchronizacja ruchu warg i mimiki twarzy ze słowami wypowiedzianymi w danym nagraniu, pozwala na utworzenie doskonałego obrazu oszustwa, umożliwiającego podszywanie się pod dowolną osobę.*” (s. 83) W ostatnim podrozdziale przybliża trzy nurty badawcze skupione na analizie: stylu w jakim powstają fałszywe wiadomości (style-based fake news analysis), ich propagacji i rozprzestrzeniania (propagation-based

fake news analysis) oraz zaangażowania użytkowników w propagowanie nieprawdziwych wiadomości (user-based fake news analysis). Rozdział kończą syntetyczne wnioski.

W rozdziale trzecim zatytułowanym „Techniczne aspekty nagrań deepfake oraz rozwój technologii" – zaprezentowano możliwości tworzenia nagrań deepfake na podstawie przeprowadzonego eksperymentu. W tym celu Doktorant własnoręcznie przygotował nagrania deepfake. Proces badawczy przeprowadzony został w kilku etapach. Pierwszy z nich obejmował organizację laboratorium badawczego poprzez skompletowanie i złożenie stacjonarnego komputera o odpowiednich parametrach pozwalających na swobodne tworzenie nagrań deepfake. W kolejnym etapie Autor dokonał przeglądu oprogramowania służącego do tworzenia nagrań deepfake (FakeApp, MyFakeApp, DeepFaceLab oraz FaceSwap), a następnie przetestował każde z nich. Do tworzenia fałszywych nagrań wykorzystał oprogramowanie DeepFaceLab (DFL). Na koniec rozdziału Doktorant sformułował uogólnienia i wnioski. Bardzo wysoko oceniam powyższy rozdział z uwagi na przeprowadzony i właściwie opisane badania empiryczne.

W rozdziale czwartym pt.: „Percepcja zmanipulowanych przekazów wizualnych w świetle procesów psychologicznych i społecznych" – przedstawiono wyniki eksperymentu pod względem zdolności badanych do rozpoznawania zmanipulowanych treści. Na początku Doktorant dokonał oceny naturalności wyglądu prezentowanych nagrań, w tym celu przeanalizował odpowiedzi na zadane respondentom pytanie, przeprowadził analizę z uwzględnieniem zmiennej nominalnej, serię testów Kruskala-Wallisa, test Friedmana oraz nieparametryczny odpowiednik testów post hoc. W dalszej kolejności przeprowadzono weryfikację prawdziwości prezentowanych nagrań oraz rozpoznawania fałszywych nagrań. Ponadto, Doktorant przeprowadził analizę odpowiedzi na fakultatywne, opisowe pytanie dotyczące wskazania nienaturalnych elementów wyglądu. Kolejna część materiału odnosi się do kwestii rozpoznania osoby występującej na nagraniu. Autor przeprowadził także badania związane z problematyką zaufania do wizerunku prezentowanej na nagraniu osoby, a także kojarzenia osoby występującej na nagraniu. Na zakończenie rozdziału Autor zawarł podsumowanie

i wnioski. Rozdział czwarty cechuje także wysoki poziom merytoryczny poparty właściwym i rzetelnym opisem przeprowadzonych badań empirycznych.

Rozdział piąty zawiera wyniki badań uzyskanych na podstawie przeprowadzonego eksperymentu, zmierzające do zidentyfikowania wpływu nagrań deepfake na bezpieczeństwo personalne. W poszczególnych podrozdziałach przedstawiono odpowiedzi respondentów na zadane pytania. W pierwszym podrozdziale Doktorant analizował odpowiedzi na pytanie dotyczące zachęcania fałszywymi nagraniami do inwestowania. W tym miejscu należy także podkreślić umiejętne wykorzystanie przez Doktoranta testu Shapiro-Wilk oraz analizy z uwzględnieniem zmiennej nominalnej E, przeprowadzenie nieparametrycznych testów Kruskal-Wallis, oraz nieparametrycznego testu Friedmana. Do opisu uzyskanych wyników badań wykorzystał narzędzie w postaci statystyki opisowej, a także histogramy (rozkłady). Podobny schemat postępowania Doktorant zastosował podczas opracowywania kolejnych podrozdziałów, analizując odpowiedzi na pytania dotyczące inwestowania własnych pieniędzy na polecanej platformie, przekonania o możliwościach osiągnięcia dużego zysku, obaw dotyczących utraty pieniędzy. W kolejnych podrozdziałach zamieszczono analizę odpowiedzi na pytania dotyczące kwestii związanych z wpływem deepfake na decyzje inwestycyjne, a także wpływu nagrania na chęć zainwestowania w dany projekt. Rozdział kończą uogólnienia i wnioski. Należy podkreślić, że Autor podczas analizy uzyskanych odpowiedzi zastosował taką samą metodologie badań jak w poprzednich rozdziałach, co pozwoliło mu właściwe opisać i zinterpretować wyniki badań, a tym samym uniknąć ich zniekształceń.

Ostatni rozdział, zatytułowany „Przeciwdziałanie dezinformacji wyzwaniem dla bezpieczeństwa strukturalnego" podzielony został na dwie części, gdzie pierwsza ukazuje wskaźniki psychologiczne (powodujące wzrost podatności na manipulację z wykorzystaniem nagrań deepfake), druga zaś sposoby przeciwdziałania ryzykom związanym z rozwojem AI.

Autor na podstawie uzyskanych odpowiedzi od respondentów na zadane pytanie próbował ustalić jak bezprawne wykorzystanie wizerunku influencera wpłynie na jego postrzeganie przez otoczenie. Dalszy materiał w zamiarze Doktoranta miał na celu

sprawdzenie w jaki sposób zachowają się respondenci po obejrzeniu każdego z przedstawionych nagrań. Kolejny podrozdział dotyczy kwestii propagacji fałszywych nagrań, a w szczególności sprawdzenia możliwości „udostępnienia" przez respondentów danego nagrania. Przedstawiona zastała również analiza korelacje moderatorów psychologicznych w odniesieniu do uzyskanych odpowiedzi na zadane pytania dotyczące: impulsywności, fundamentów moralnych, potrzeb poznawczego domknięcia, a także moderatorów samooceny, depresji i lęku. Na koniec zaprezentowano autorskie rozwiązania zmierzające do zwiększenia świadomości społecznej na zagrożenia związane z technologią deepfake.

W zakończeniu rozprawy Autor ocenia stopień osiągnięcia zakładanego celu głównego, rozwiązania problemu badawczego i stopień zweryfikowania hipotezy głównej i hipotez szczegółowych. Część merytoryczną pracy zamyka bibliografia i co należy wyartykułować Doktorant trafnie dobrał literaturę przedmiotu.

Ocena rozprawy

Warunki stawiane rozprawom doktorskim, a zarazem cechy, jakie powinna wykazywać przedmiotowa dysertacja zostały określone w art. 187 ustawy z dnia 20 lipca 2018 r. *Prawo o szkolnictwie wyższym i nauce*. Ponadto, zgodnie z zapisami art. 185 zawartymi w poradniku Rady Doskonałości Naukowej dla recenzentów (2022 r.) opinia dotycząca danej rozprawy doktorskiej powinna zawierać następujące elementy:

1) ocenę wraz z uzasadnieniem, czy rozprawa doktorska prezentuje ogólną wiedzę teoretyczną osoby ubiegającej się o nadanie stopnia doktora w określonej dyscyplinie albo dyscyplinach;
2) ocenę wraz z uzasadnieniem, czy rozprawa doktorska wykazuje umiejętność samodzielnego prowadzenia pracy naukowej (…) przez osobę ubiegającą się o nadanie stopnia doktora;
3) ocenę wraz z uzasadnieniem, czy rozprawa doktorska stanowi oryginalne rozwiązanie problemu naukowego, oryginalne rozwiązanie w zakresie zastosowania wyników własnych badań naukowych w sferze gospodarczej lub społecznej (…).

Mając powyższe na względzie, przedstawiam moją opinię:

Ad. 1 Stwierdzam, że Doktorant w rozprawie prezentuje w stopniu wystarczającym ogólną wiedzę teoretyczną w dyscyplinie nauki o bezpieczeństwie. Przedstawiam uzasadnienie powyższego stwierdzenia:

a) Doktorant wiedzę w zakresie przedmiotu badań oraz nauk o bezpieczeństwie zdobywał w czasie prowadzonych badań, a w szczególności podczas prowadzonej analizy danych zastanych. Ponadto, z obszaru poruszanej problematyki prowadził szkolenia specjalistyczne, a także zajęcia na Uniwersytecie Warszawskim. Swoją wiedzą dzielił się także podczas organizowanych ogólnopolskich oraz międzynarodowych konferencji naukowych.
b) Autor w badaniach wielokrotnie odwołuje się do opracowań teoretyków i praktyków w dziedzinie bezpieczeństwa, w tym bezpieczeństwa informacyjnego (zob. Bibliografia). Korzystając z literatury przedmiotu (o czym świadczą liczne cytaty oraz przepisy), demonstruje umiejętność efektywnego wykorzystania zdobytej wiedzy.
c) Rozprawa dowodzi, że Doktorant wypracował zdolność dotarcia do unikatowych źródeł informacji i wiedzy często zawartych w opracowaniach anglojęzycznych. Jak podkreśla Autor *"Z racji aktualności tematu, jego świeżości oraz dzięki szybkiemu rozwojowi, w nauce brakuje zwartych publikacji poświęconych tej problematyce"*. (s. 19-20), dlatego też należy docenić wysiłek Autora zmierzający do zaprezentowania dorobku aktualnego i wysoce wartościowego.

Ad.2. Stwierdzam, że rozprawa doktorska wykazuje umiejętności Doktoranta do prowadzenia pracy naukowej. Przedstawiam uzasadnienie powyższego stwierdzenia:

a) Doktorant opanował umiejętność poprawnego formułowania założeń badawczych. Prawidłowo określił główny cel badań oraz cel utylitarny. Dopełnieniem rozważań naukowego wywodu jest właściwie sformułowany główny problem badawczy w postaci pytania: *Czy technologia deepfake pozwala na generowanie rzeczywistych nagrań, mogących wpływać na decyzje osób je oglądających, a przez to zagrażać bezpieczeństwu narodowemu?* W pełni koresponduje z nim główna hipoteza robocza.

W rozprawie sformułowano sześć problemów szczegółowych oraz adekwatnie do nich przyjęto hipotezy robocze.

b) Doktorant wykazał się bardzo dobrą znajomością różnych metod badawczych i umiejętnością ich wykorzystania do realizacji postawionych przed sobą celów badawczych. Pan Bartosz Biderman oprócz metod teoretycznych wykorzystał również metody empiryczne, takie jak: analiza dokumentów oraz eksperyment badawczy. Przedstawione interpretacje otrzymanych wyników badań są bardzo dokładne i pogłębione, a rysunki bardzo przejrzyste. Należy podkreślić, że bardzo wysoko oceniam wyniki badań empirycznych oraz właściwą ich interpretację w formie opisu. Szczególnie imponujące są wyniki badań przeprowadzonego eksperymentu, które Doktorant znakomicie wykorzystał podczas opracowywania rozdziałów merytorycznych.

c) Lektura rozprawy doktorskiej wskazuje na bardzo dobre przygotowanie Doktoranta do samodzielnego prowadzenia badań naukowych. Umiejętnie korzysta z materiału badawczego, poddaje go analizie realizując przyjęte założenia wynikające z przyjętego procesu badawczego.

Ad. 3. Stwierdzam, że rozprawa doktorska stanowi oryginalne rozwiązanie problemu naukowego, oryginalne rozwiązanie w zakresie zastosowania wyników własnych badań naukowych w sferze społecznej. Przedstawiam uzasadnienie powyższego stwierdzenia:

a) Pan Bartosz Biderman podjął się ambitnego zadania, polegającego na zidentyfikowaniu technologii deepfake oraz metod manipulowania obrazem i dźwiękiem, a także wskazaniu działań zmierzających do przeciwdziałania tym zagrożeniom. Zdołał właściwie zastosować badania teoretyczne i empiryczne, co pozwoliło na wydobycie unikatowych wyników badań.

b) Doktorant wnikliwie zidentyfikował i wypełnił lukę w dotychczasowej nauce dotyczącej szeroko pojmowanego bezpieczeństwa informacyjnego. Autor zwrócił uwagę na dynamicznie zmieniające się zagrożenia i uwarunkowania związane z rozwojem technologii deepfake.

c) Temat rozprawy, jak i podjęte w niej problemy badawcze są istotne, szczególnie w obliczu współczesnych zagrożeń związanych z dynamicznym

rozwojem technologii deepfake i mieszczą się w głównym nurcie ważnej i aktualnej tematyki badań nauk o bezpieczeństwie. Uważam zatem, za w pełni uzasadnione podjęcie tematu recenzowanej monografii, przeprowadzenie badań i ich pisarskie opracowanie w formie rozprawy doktorskiej.

Uwagi krytyczne

W podrozdziale 1.1.2. *Przedmiot i cel badań* - w mojej opinii brakło właściwego uwypuklenia (wyartykułowania) przedmiotu badań. Przedmiot badań powinien być wyraźnie sformułowany i wskazany w tekście np. „przedmiotem badań w niniejszym opracowaniu jest...". Pewne moje wątpliwości budzi też przyjęty podział celu badawczego: na cel główny oraz cel praktyczny (utylitarny). Zazwyczaj w literaturze przedmiotu wskazuje się, że cele mogą przyjmować następujące formy i rodzaje: a) ze względu na zadania, jakie mają realizować: cel teoretyczny, cel poznawczy oraz cel praktyczny; b) ze względu na poziom szczegółowości: cel główny oraz cele cząstkowe.

W odniesieniu do problemów szczegółowych kwestią do wyjaśnienia jest sformułowanie przez Autora dwóch celi do rozdziału szóstego, a do wcześniejszych rozdziałów po jednym.

Nie do końca przekonują mnie także sformułowane problemy szczegółowe do rozdziału 3 i 4 (*w jaki sposób powstają nagrania deepfake*?, *Czy internauci rozróżniają materiały multimedialne prawdziwe od fałszywych*?), które mają bardziej charakter ogólny niż naukowy, albowiem nie wskazują na problem który należy rozwiązać.

W rozprawie zauważalny jest brak zachowania dysproporcji między rozdziałami Niektóre rozdziały liczą trzy, cztery podrozdziały (np. rozdział 3 i 4), a inne sześć, siedem, osiem (np. rozdział 4, 5, 6). Ponadto, materiał zamieszczony w rozdziale szóstym a w szczególności podrozdziały: 6.2, 6.3 i 6.4 w mojej ocenie nie do końca wpisuje się w tytuł rozdziału pt. „Przeciwdziałania dezinformacji wyzwaniem dla bezpieczeństwa strukturalnego". Może lepiej byłoby zamieścić ten materiał w rozdziale 5, a w tym miejscu więcej uwagi poświecić problematyce np. oceny dotychczasowych działań podejmowanych w celu przeciwdziałania dezinformacji z perspektywy bezpieczeństwa narodowego, czy też zaproponowanie zmian

w przepisach prawnych i innych obszarach, ale oczywiście jest to kwestia dyskusyjna i moja subiektywna ocena.

Mimo pewnych uwag z dużym uznaniem odnoszę się do wartości metodycznej dysertacji podkreślając dojrzałość Autora do samodzielnego prowadzenia badań. Metodologiczną część rozprawy oceniam pozytywnie, a wskazane wątpliwości nie rzutują na ogólną wartość przyjętych założeń.

Pytania

1) Czy może Pan doprecyzować przyjęty w rozprawie przedmiot badań?
2) Jaką metodologię przyjął Pan przy formułowaniu celu badawczego, dzieląc go na cel główny oraz praktyczny?
3) Proszę uzasadnić, dlaczego do rozdziału szóstego sformułował Pan dwa problemy szczegółowe, a do pozostałych po jednym?
4) Czy może Pan w kilku punktach wymienić propozycje przeciwdziałania dezinformacji multimedialnej z wykorzystaniem technologii deepfake w ocenie Pana kluczowe dla systemu bezpieczeństwa narodowego?

Wnioski końcowe

Treść rozprawy wskazuje, że Doktorant posiada gruntowną wiedzę teoretyczno-praktyczną w obszarze bezpieczeństwa, a także ukazuje umiejętność samodzielnego prowadzenia pracy naukowej. Przedstawiona do recenzji dysertacja spełnia wymagania stawiane rozprawom doktorskim z dziedziny nauk społecznych, w dyscyplinie nauki o bezpieczeństwie, zgodnie z ustawą z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (t.j. Dz.U. z 2024 r. poz. 1571 z późn. zm.) art. . 187.

W związku z powyższym wnoszę o dopuszczenie Doktoranta do publicznej obrony przedłożonej rozprawy doktorskiej.

Bogdan Grenda