



dr hab. inż. Piotr A. Kowalski, prof. AGH  
Katedra Informatyki Stosowanej i Fizyki Komputerowej,  
Wydział Fizyki i Informatyki Stosowanej,  
Akademia Górniczo-Hutnicza w Krakowie,  
al. Mickiewicza 30, 30-059 Kraków  
email: [pkowal@agh.edu.pl](mailto:pkowal@agh.edu.pl)



oraz

Centrum Informatycznych Metod Analizy Danych  
Instytut Badań Systemowych  
Polskiej Akademii Nauk

Kraków, 25.05.2026

## RECENZJA

### **rozprawy doktorskiej Pana mgr. inż. Witolda Sosnowskiego pt. „Classifying and Mining Disinformation Narratives With Large Language Models”**

#### **1. Uwagi ogólne**

Rozprawa doktorska mgr. inż. Witolda Sosnowskiego pt. „Classifying and Mining Disinformation Narratives With Large Language Models”, przygotowana pod opieką promotorską prof. dr. hab. Adama Wierzbickiego oraz promotor pomocniczej dr Kingi Skorupskiej, podejmuje temat wyjątkowo aktualny, społecznie ważny i naukowo wymagający. Autor koncentruje się na problemie klasyfikowania oraz automatycznego wydobywania narracji dezinformacyjnych z wykorzystaniem dużych modeli językowych. Już sam wybór przedmiotu badań zasługuje na uznanie, ponieważ dotyczy obszaru, w którym rozwój technologii sztucznej inteligencji spotyka się bezpośrednio z bezpieczeństwem informacyjnym, jakością debaty publicznej oraz odpornością instytucji demokratycznych.

Rozprawa została napisana w języku angielskim. Jej struktura jest przejrzysta, logiczna i dobrze podporządkowana celom badawczym. Praca składa się z abstraktu, streszczenia w języku polskim, siedmiu zasadniczych rozdziałów, bibliografii oraz trzech obszernych dodatków zawierających między innymi szczegóły dotyczące benchmarku EUDisinfoTest, metody DiNaM oraz metody DiNO. Układ pracy prowadzi od motywacji i przeglądu literatury, poprzez definicje i formalizację zadań badawczych, aż po trzy zasadnicze komponenty empiryczno-metodyczne. Po rozdziałach wprowadzających i przeglądowych Autor przechodzi do definicji i sformułowań zadań, następnie przedstawia benchmark EUDisinfoTest, metodę DiNaM, metodę DiNO, a całość zamyka rozdziałem podsumowującym wraz z ograniczeniami i kierunkami dalszych badań.

Bibliografia jest obszerna i adekwatna do interdyscyplinarnego charakteru pracy. Obejmuje literaturę z zakresu dezinformacji, fact-checkingu, komunikacji politycznej, analizy narracji, przetwarzania języka naturalnego oraz dużych modeli językowych. Widać, że Autor swobodnie porusza się zarówno w obszarze technologicznym, jak i w szerszym kontekście nauk społecznych. Ważnym elementem pracy są także dodatki, które dokumentują szczegóły promptów, procedur ewaluacyjnych, wytycznych anotacyjnych i przykładów, wzmacniając reprodukowalność badań. Rozprawa nie formułuje jednej, odrębnie wyodrębnionej tezy w postaci klasycznego zdania typu „tezą pracy jest...”. Jej zasadnicza teza wynika jednak jasno z całego wywodu: współczesna dezinformacja powinna być analizowana nie tylko na poziomie pojedynczych twierdzeń, artykułów czy postów, lecz przede wszystkim na poziomie powtarzalnych narracji, a duże modele językowe — odpowiednio oceniane, kontrolowane i zakotwiczone w danych fact-checkingowych — mogą wspierać zarówno klasyfikację, jak i wydobywanie takich narracji. Autor sam wskazuje, że celem rozprawy jest przejście od klasyfikacji dezinformacji na poziomie pojedynczych elementów do rozumienia narracyjnego, obejmującego ewaluację, mapowanie globalne oraz analizę na poziomie konkretnych źródeł medialnych.

Praca doktorska posiada trójdzielny charakter. Po pierwsze, Autor wprowadza EUDisinfoTest — benchmark służący do oceny zdolności dużych modeli językowych do rozróżniania narracji dezinformacyjnych od wiarygodnych, także w wariantach retorycznych Logos, Ethos i Pathos. Po drugie, proponuje DiNaM — skalowalną metodę wydobywania globalnych map narracji dezinformacyjnych z dużych zbiorów artykułów fact-checkingowych. Po trzecie, przedstawia DiNO — metodę ukierunkowaną na analizę narracji dezinformacyjnych w konkretnych źródłach medialnych, z wykorzystaniem baz zweryfikowanych fałszywych twierdzeń.

Za najważniejsze osiągnięcia rozprawy uznaję następujące elementy. Po pierwsze, Autor trafnie identyfikuje lukę pomiędzy praktyką monitorowania dezinformacji, która coraz częściej operuje kategorią narracji, a istniejącymi metodami obliczeniowymi, często skupionymi na pojedynczych twierdzeniach. Po drugie, rozprawa wprowadza uporządkowaną formalizację dwóch zadań: klasyfikacji narracji dezinformacyjnych oraz wydobywania takich narracji z korpusów tekstowych. Po trzecie, EUDisinfoTest stanowi wartościowy wkład w ewaluację LLM-ów, ponieważ nie ogranicza się do prostego rozpoznawania prawdziwości, lecz bada podatność modeli na retoryczne przeformułowanie tej samej treści. Po czwarte, DiNaM oferuje praktyczną i empirycznie sprawdzoną ścieżkę od artykułów fact-checkingowych do globalnej mapy narracji. Po piąte, DiNO uzupełnia tę perspektywę o poziom konkretnych źródeł medialnych, co jest istotne zarówno badawczo, jak i aplikacyjnie. Po szóste, praca opiera się na znaczącym dorobku publikacyjnym, w tym dwóch publikacjach EMNLP oraz pracy rozwijanej dalej w kierunku DiNO.

Oceniam rozprawę bardzo wysoko. Jest to praca dojrzała, ambitna, nowoczesna i dobrze osadzona w aktualnych problemach badawczych. Jej wartość polega nie tylko na zaproponowaniu konkretnych algorytmów, lecz także na zmianie perspektywy: od punktowego wykrywania fałszywych informacji do rozumienia szerszych, powtarzalnych struktur narracyjnych. Taka zmiana jest ważna, ponieważ dezinformacja rzadko funkcjonuje jako pojedyncze, odizolowane zdanie; częściej przyjmuje postać powracających opowieści, które adaptują się do nowych kontekstów, tematów i odbiorców.

## **2. Ocena rozprawy**

### **a. Uwagi krytyczno-polemiczne:**

W niniejszej części pokrótce przedstawione zostaną główne mankamenty dysertacji. Ze względu na istotność tejże części recenzji niniejsze uwagi zostaną przedstawione w punktach, do których łatwiej będzie się odnieść Doktorantowi.

1. Wyraźniejsze sformułowanie tezy głównej. Chociaż cele badawcze O1–O4 są jasno przedstawione, warto byłoby w części wprowadzającej sformułować jedną syntetyczną tezę rozprawy. Obecnie, czytając dysertację, można rekonstruować ją z motywacji, celów i wniosków. Nie jest to błąd merytoryczny, ponieważ konstrukcja pracy pozostaje czytelna, ale jedno zdanie tezy głównej mogłoby dodatkowo wzmocnić kompozycję rozprawy.
2. Autor przyjmuje definicję dezinformacji obejmującą fałszywość, potencjalną szkodliwość oraz intencję wprowadzenia w błąd. W praktyce jednak algorytmiczne ujęcie pracy opiera się głównie na zweryfikowanej fałszywości i zasobach fact-checkingowych. Warto byłoby szerzej przedyskutować, w jakim stopniu proponowane metody rzeczywiście uchwytują intencję, a w jakim pozostają narzędziami do mapowania fałszywych lub wprowadzających w błąd narracji.
3. Rozprawa dobrze rozróżnia narracje, ramy interpretacyjne i strategie retoryczne, ale w zastosowaniach praktycznych granice te bywają płynne. Szczególnie w DiNO, gdzie analizowane są konkretne źródła medialne, warto byłoby jeszcze mocniej pokazać, kiedy wydobyty wzorzec jest już narracją dezinformacyjną, a kiedy jedynie powtarzalnym framingiem, opinią lub nacechowaną interpretacją.
4. EUDisinfoTest jest jednym z najmocniejszych elementów pracy. Jednocześnie kontrolowane warianty Logos, Ethos i Pathos są tworzone z udziałem LLM-ów i następnie weryfikowane przez ekspertów. Warto byłoby rozwinąć dyskusję o tym, czy przeformułowanie retoryczne zawsze zachowuje dokładnie tę samą treść propozycyjną, czy też czasem może wzmocniać, osłabiać lub przesuwać sens narracji. To szczególnie ważne, ponieważ wyniki pokazują, że modele są wrażliwe na retoryczny sposób przedstawienia treści.
5. W kilku komponentach najlepsze wyniki osiągają modele zamknięte. Jest to zrozumiałe z punktu widzenia jakości, ale rodzi pytania o reprodukowalność, audytowalność i stabilność

wyników w dłuższym horyzoncie. Modele komercyjne zmieniają się, a ich dokładne dane treningowe i mechanizmy działania są niejawne. Autor zauważa ten problem w ograniczeniach pracy; warto byłoby jeszcze silniej powiązać go z praktycznym wdrożeniem systemów monitorowania dezinformacji.

6. W pracy zastosowano metryki semantyczne, takie jak Weighted Chamfer Distance, Silhouette Score oraz oceny zgodności tematycznej i zgodności stanowiska. Są to rozwiązania uzasadnione i praktyczne, ale narracja jest konstruktem złożonym, a jej jakość nie zawsze w pełni daje się sprowadzić do podobieństwa embeddingów lub spójności klastra. Dobrym uzupełnieniem byłaby szersza ewaluacja ekspercka, obejmująca nie tylko zgodność z referencją, lecz także użyteczność narracji dla analityków.
7. Analizy temporalne DiNaM i DiNO są bardzo interesujące, ponieważ pokazują, jak określone narracje nasilają się wokół ważnych wydarzeń społeczno-politycznych. Autor ostrożnie zaznacza, że są to zależności korelacyjne. W dalszych badaniach warto byłoby rozważyć bardziej formalne podejście do testowania zmian w czasie, np. modele wykrywania punktów zwrotnych, analizy przyczynowości w sensie statystycznym lub badanie opóźnień między wydarzeniami a aktywacją narracji.

W szczególności z racji na obowiązki recenzenta, muszę w tym miejscu wskazać jeszcze kilka uwag związanych z pojedynczymi niedociągnięciami stylistycznymi. Doktorant również nie ustrzegł się nielicznych mankamentów natury technicznej, takich jak błędy interpunkcyjne, czy też w pracy występują nieliczne usterki edytorskie np. odwołania typu „?” oraz literówki w opisach rysunków itp., jednak w dużej mierze nie rzutują one na ocenę pracy.

Poniższe uwagi mają charakter porządkujący i nie podważają wartości pracy.

1. Rozdział 1.3, s. 4–6 — niespójność numeracji wkładów. Autor pisze, że rozprawa wnosi dziewięć wkładów oznaczonych jako C1–C9, natomiast w widocznej części rozdziału przedstawione są wkłady C1–C7, a tabela mapująca cele, rozdziały i wkłady również obejmuje C1–C7. Należy albo dopisać brakujące C8–C9, albo zmienić sformułowanie „nine contributions (C1–C9)” na zgodne z faktyczną listą.
2. Rozdział 1.5, s. 10 — niezręczność językowa. W opisie rozdziału 5 występuje fraza „evaluates the method analyzing alignment”. Poprawniej byłoby: „evaluates the method by analysing alignment” albo „evaluates the method through an analysis of alignment”.
3. Rozdział 3.2.1, s. 26 — literówka. W zdaniu „based on the ddefinitions of disinformation narratives...” występuje literówka „ddefinitions”. Należy poprawić na „definitions”.
4. Rozdział 4.7.3, s. 42 — powtórzony akapit i niespójność liczby. Dwukrotnie występuje akapit zaczynający się od „In the Base condition, models already perform strongly...”. W pierwszej wersji wskazano średni TPR 0.90, w drugiej 0.89.

5. Rozdział 5.1, s. 47 — powtórzenie treści. Dwa następujące po sobie akapity zaczynają się od „Within the overall research programme...” i pełnią bardzo podobną funkcję kompozycyjną. Warto je połączyć lub skrócić, aby uniknąć wrażenia redakcyjnej duplikacji.
6. Rozdział 6.3.3, s. 83 — brakujące odwołanie bibliograficzne. Przy charakterystyce The Guardian występuje „[?]” w zdaniu dotyczącym wysokiej oceny rzetelności i wiarygodności redakcyjnej. Należy uzupełnić właściwy przypis bibliograficzny.
7. Rozdział 7.2.2, s. 117 — interpunkcja i styl nagłówek. Nagłówek „Media Science, and Political Science” zawiera zbędny przecinek. Naturalniejsza forma to „Media Science and Political Science” albo „Media and Political Science”.
8. Rozdział 7.5, s. 121 — błąd gramatyczny. W sformułowaniu „LLM-based narratives classification” należałoby użyć „LLM-based narrative classification” albo „LLM-based classification of narratives”.

### **3. Podsumowanie**

Przedstawiona rozprawa doktorska jest wartościowym i dojrzałym osiągnięciem naukowym. Autor podejmuje problem o dużym znaczeniu społecznym i technologicznym, a jednocześnie proponuje konkretne, empirycznie zweryfikowane rozwiązania. Na szczególne uznanie zasługuje połączenie perspektywy technicznej z refleksją nad praktyką fact-checkingu i analizą komunikacji publicznej. Rozprawa nie ogranicza się do zastosowania dużych modeli językowych jako gotowych narzędzi, lecz bada ich ograniczenia, podatność na retorykę oraz przydatność w złożonych zadaniach wydobywania narracji.

Uwagi przedstawione powyżej mają przede wszystkim charakter dyskusyjny i porządkujący. Wskazują możliwe kierunki dalszego rozwoju tematu, doprecyzowania metodologii oraz redakcyjnego ulepszenia tekstu. Nie wpływają one na moją jednoznacznie pozytywną ocenę rozprawy. Praca stanowi oryginalny wkład w obszar przetwarzania języka naturalnego, badań nad dezinformacją oraz zastosowań dużych modeli językowych w analizie narracji.

**Konkludując uważam, że rozprawa doktorska Pana mgr. inż. Witolda Sosnowskiego spełnia wymagania stawiane w odpowiednich przepisach rozprawom doktorskim i wobec tego stawiam wniosek o jej dopuszczenie do dalszych, przewidzianych Ustawą, etapów postępowania o nadanie stopnia doktora.**

W związku z bardzo wysoką wartością naukową rozprawy, jej wyjątkowo aktualną i społecznie istotną tematyką, a także znaczącym dorobkiem publikacyjnym Doktoranta bezpośrednio powiązanym z prowadzonymi w dysertacji badaniami, zwracam się do Wysokiej Rady Naukowej ITiT w Polsko-Japońskiej Akademii Technik Komputerowych, z wnioskiem o wyróżnienie rozprawy doktorskiej mgr. inż. Witolda Sosnowskiego. W mojej ocenie praca ta stanowi dojrzałe, oryginalne i ważne osiągnięcie naukowe, które nie tylko wpisuje się w najnowsze kierunki badań nad dużymi modelami językowymi i dezinformacją, lecz także wnosi istotny wkład w rozwój metod analizy narracji dezinformacyjnych.

dr hab. inż. Piotr A. Kowalski, prof. AGH