**РЕЦЕНЗІЯ**

на дисертаційну роботу

**Падалко Галини Анатоліївни**

на тему «Моделі, методи та інформаційна технологія виявлення і аналізу текстової дезінформації та пропаганди в соціальних мережах»,

яка представлена на здобуття ступеня доктора філософії

в галузі знань 12 Інформаційні технології за спеціальністю 122 Комп’ютерні науки

**Актуальність теми.** Сьогодні дезінформація та пропагандистські наративи перетворилися на основні інструменти інформаційного впливу. Особливу динаміку цьому процесу надають соціальні мережі, які сприяють швидкому розповсюдженню маніпулятивного контенту та ускладнюють виявлення джерел інформаційних атак. Російський уряд систематично використовує інформаційні кампанії для підриву довіри до державних інституцій, провокування внутрішніх конфліктів та дестабілізації суспільства. Такі дії неодноразово фіксувалися у звітах міжнародних організацій та аналітичних центрів. У зв’язку з цим зростає потреба у впровадженні інструментів на базі штучного інтелекту, здатних оперативно обробляти великі масиви даних і виявляти дезінформаційні мережі. Актуальність дослідження зумовлена необхідністю зміцнення інформаційної безпеки та посилення стійкості суспільства до зовнішніх інформаційних загроз.

**Оцінка обґрунтованості наукових результатів дисертації, їх достовірності та новизни.** Матеріали, представлені у дисертаційній роботі Падалко Галини Анатоліївни, свідчать про наукову обґрунтованість і достовірність отриманих результатів. Це забезпечується застосуванням сучасних дослідницьких підходів, зокрема заснованих на теоріях обчислювального штучного інтелекту – машинного та глибокого навчання, що дозволяють ефективно класифікувати й кластеризувати великі обсяги даних. У результаті проведеного аналізу в дисертаційній роботі сформульовано та вирішено актуальне наукове завдання – розробку технології дослідження текстових даних з метою виявлення інформаційних маніпуляцій.

Розроблені авторкою моделі, методи та програмні рішення впроваджені в діяльність Balsillie School of International Affairs (акт впровадження від 18 березня 2025 р.), Center for International Governance Innovation (акт впровадження від 18 березня 2025 р.), Державної установи "Харківський обласний центр контролю та профілактики хвороб МОЗ України" (акт впровадження від 2 грудня 2024 р.), а також у навчальний процес (акт впровадження від 30 січня 2025 р.) і наукову діяльність (акт впровадження від 4 березня 2025 р.) Національного аерокосмічного університету «Харківський авіаційний інститут».

**Наукова новизна** результатів дисертаційного дослідження полягає в наступному:

1. Вперше запропоновано фреймворк комплексного аналізу текстової дезінформації та пропаганди, заснований на комбінуванні технологій двоспрямованого кодувального представлення з трансформерів (BERT) та двоспрямованої довгої короткочасної пам’яті (Bi-LSTM) для створення інтерпретованих векторних представлень тексту, який на відміну від існуючих технік, об’єднує виявлення та тематичний аналіз дезінформації, що дозволяє забезпечити ефективне виявлення та аналіз текстової дезінформації та пропаганди в соціальних мережах.

2. Вперше розроблено модель класифікації текстової дезінформації та пропаганди, заснований на поєднанні трансформерів і архітектури Bi-LSTM для інтеграції лінійних та нелінійних залежностей текстових даних, яка на відміну від існуючих використовує глибокі контекстуальні ембединги і механізми уваги, що дозволяє підвищити точність моделі.

3. Удосконалено моделі глибокого навчання для класифікації текстової дезінформації та пропаганди, засновані на архітектурах XLNet, Bi-LSTM та Attention-Based Bi-LSTM, які на відміну від існуючих застосовують адаптивне оцінювання важливих фрагментів тексту, що дозволяє забезпечувати високий рівень продуктивності класифікації.

4. Удосконалено метод тематичного моделювання, заснований на застосуванні механізмів багатоголової уваги, що на відміну від існуючих використовує глибокі ембединги для контекстуалізації даних, що дозволяє забезпечити чітке розмежування між кластерами.

5. Дістали подальшого розвитку моделі для класифікації текстової дезінформації та пропаганди, засновані на ансамблевих методах машинного навчання, які на відміну від існуючих враховують балансування класів і адаптуються до змін вхідних даних, що дозволяє підвищити продуктивність моделей.

6. Дістали подальшого розвитку моделі для класифікації інформації, засновані на статистичних методах машинного навчання, які на відміну від існуючих адаптовані до класифікації текстової дезінформації та пропаганди, що дозволяє створення систем для виявлення місінформації, дезінформації та малінформації.

**Оцінка змісту дисертації, її завершеність та дотримання принципів академічної доброчесності, мова та стиль викладення результатів.** Зміст дисертаційної роботи Падалко Галини Анатоліївни повністю відповідає вимогам Стандарту вищої освіти за спеціальністю 122 «Комп’ютерні науки». Дисертаційна робота є завершеним науковим дослідженням, результати якого свідчать про суттєвий особистий внесок авторки у розвиток відповідного наукового напряму.

Дисертаційна робота є результатом самостійної наукової діяльності здобувачки та не містить ознак фальсифікації, компіляції, фабрикації, плагіату або неналежного запозичення. Усі використані ідеї, результати досліджень та фрагменти текстів інших авторів коректно оформлені з відповідними бібліографічними посиланнями на джерела.

Дисертаційна робота виконана українською мовою з дотриманням норм наукового стилю, із використанням коректної термінології, властивої фаховим і науковим дослідженням у відповідній галузі. Робота відзначається цілісністю викладу, повнотою охоплення теми та чітким розкриттям основної наукової ідеї авторки. Представлені у дисертаційній роботі теоретичні та прикладні положення викладені послідовно, логічно та належно обґрунтовані.

Структура дисертаційної роботи охоплює анотацію, зміст, перелік умовних скорочень, вступ, чотири розділи основної частини, висновки, список використаних джерел та додатки. Загальний обсяг роботи становить 207 сторінок друкованого тексту, зокрема: анотація – 7 сторінок, зміст – 4 сторінки, перелік умовних скорочень – 2 сторінки, основна частина – 124 сторінки, список із 151 джерела – 17 сторінок, додатки – 22 сторінки. Робота містить 54 рисунки та 16 таблиць.

**Зміст дисертаці**ї. У вступі обґрунтовано актуальність теми, сформульовано мету, завдання, об’єкт і предмет дослідження, окреслено наукову новизну, методологічні підходи, ступінь апробації та зв’язок із науковими програмами.

**Розділ 1** присвячений теоретичному аналізу дезінформації, місінформації та малінформації в умовах гібридної війни, з акцентом на їх використання державними акторами через соціальні мережі й штучний інтелект. Описано сучасні методи виявлення інформаційних загроз.

**Розділ 2** містить порівняльний аналіз алгоритмів машинного навчання (SVM, LightGBM тощо) для виявлення дезінформації. Найкращі результати продемонстровані на корпусах WELFake і Weibo, із точністю до 89% та F1-мірою 0.977.

**Розділ 3** присвячено глибоким моделям (BERT, XLNet, BiLSTM тощо) та запропонованій гібридній архітектурі з attention-механізмом, яка досягла точності 99.47% на різних датасетах.

**Розділ 4** описує інтегрований фреймворк, що поєднує класифікацію та тематичне моделювання (BERTopic) для виявлення ключових наративів. Модель виявляє ідеологічні патерни з високою точністю (до 99.47%).

**У висновках** підсумовано наукові й практичні результати, окреслено напрями подальших досліджень.

**Додатки** включають список публікацій, дані про апробацію, акти впровадження та опис використаних датасетів.

Дисертаційна робота оформлена відповідно до вимог, сформульованих у наказі МОН України від 12 січня 2017р. № 40 «Про затвердження вимог до оформлення дисертації».

**Повнота викладення результатів досліджень в опублікованих працях.** Дослідження, результати яких викладено в дисертаційній роботі, виконано на кафедрі математичного моделювання та штучного інтелекту Національного аерокосмічного університету «Харківський авіаційний інститут» в рамках виконання науково-дослідних робіт за проєктами «Розробка методологічного та інструментального забезпечення Agile трансформації процесів відбудови медичних закладів України для подолання розладів здоров’я населення у воєнний та повоєнний періоди» (Національний фонд досліджень України, № 2022.01/0017), «Uncovering Information Warfare: Detecting Russian Propaganda on Social Media» (MITACS, №IT37637), «Understanding and Mitigating the Risks of Generative AI in Propaganda and Informational Warfare» (MITACS and Center for International Governance Innovation, №IT36431).

Основні положення, ідеї, висновки дисертаційної роботи доповідалися та обговорювалися на 4th International Workshop of IT-Professionals on Artificial Intelligence, ProfIT AI 2024 (Кембрідж, США, 2024), V Міжнародній науково-практичній конференції ІТ-професіоналів та аналітиків комп’ютерних систем “ProfIT Conference” (Харків, 2023), 2023 13th International Conference on Dependable Systems, Services and Technologies, DESSERT 2023 (Афіни, Греція, 2023), 3rd International Workshop of IT-Professionals on Artificial Intelligence, ProfIT AI 2023 (Ватерлу, Канада, 2023), 17th World Congress on Public Health (Рим, Італія, 2023), IV Міжнародній науково-практичній конференції ІТ-професіоналів та аналітиків комп’ютерних систем «ProfIT Conference» (Харків, 2021), ІІІ Міжнародній науково-практичній конференції ІТ-професіоналів та аналітиків комп’ютерних систем «ProfIT Conference» (Харків, 2020).

**Недоліки та зауваження до дисертаційної роботи:**

1. Опис попередньої обробки твітів згадує видалення URL, хештегів і стоп-слів, але конкретний список стоп-слів, регекс-патернів чи використовуваний стемер не вказано.

2. У таблицях 2.1–2.4 продуктивність класичних моделей (LR, NB, KNN, SVM) наведено лише до 5 000 твітів хоча загальний корпус має 20 000+ записів.

3. У підрозділі 2.3.3 перелічені діапазони регуляризації lambda\_l1 та lambda\_l2 для LightGBM (1E-8 … 10) і згадано «найкращі гіперпараметри», проте не вказано фінальні підібрані значення.

4. Робота детально описує точність, повноту та міру F1, але не торкається питань упередженості (fairness) чи диференційної приватності, які дедалі важливіші в контексті публічних політичних даних.

5. В дисертаційній роботі наявні деякі граматичні та стилістичні помилки.

Висновок про дисертаційну роботу. Вважаю, що дисертаційна робота здобувачки ступеня доктора філософії Падалко Галини Анатоліївни на тему «Моделі, методи та інформаційна технологія виявлення і аналізу текстової дезінформації та пропаганди в соціальних мережах» виконана на високому науковому рівні, не порушує принципів академічної доброчесності та є закінченим науковим дослідженням, сукупність теоретичних та практичних результатів якого розв'язує наукове завдання, що має істотне значення для галузі знань 12 Інформаційні технології. Дисертаційна робота за актуальністю, практичною цінністю та науковою новизною повністю відповідає вимогам чинного законодавства України, що передбачені в Наказі МОН України Nº40 від 12.01.2017 «Про затвердження вимог щодо оформлення дисертації (зі змінами)» та «Порядку присудження ступеня доктора філософії та скасування рішення разової спеціалізованої вченої ради закладу вищої освіти, наукової установи про присудження ступеня доктора філософії», затвердженого Постановою КМУ від 12.01.2022 Nº44.

Здобувачка Падалко Галина Анатоліївна заслуговує на присудження ступеня доктора філософії в галузі знань 12 Інформаційні технології за спеціальністю 122 Комп’ютерні науки.

Рецензент:

доктора технічних наук,

професор, професор кафедри

комп'ютерних систем, мереж і

кібербезпеки

Національного аерокосмічного університету   
«Харківський авіаційний інститут» Ольга МОРОЗОВА