

ВІДГУК
офіційного опонента
на дисертаційну роботу
Падалко Галини Анатоліївни
на тему “Моделі, методи та інформаційна технологія виявлення і
аналізу текстової дезінформації та пропаганди в соціальних мережах”
представлену на здобуття ступеня доктора філософії
в галузі знань 12 Інформаційні технології
за спеціальністю 122 Комп’ютерні науки

Актуальність теми. У сучасних умовах гібридної війни, з якою стикається Україна, дезінформація та пропаганда з боку агресора стали ключовими інструментами інформаційного впливу. Нині дезінформація оперативно поширюється через соціальні мережі, що сприяє швидкому розповсюдженню маніпулятивного контенту та ускладнює ідентифікацію джерел атак. Росія як агресор активно застосовує інформаційні операції для дискредитації державних інституцій, поляризації суспільства та підриву довіри до офіційної інформації, що підтверджується численними дослідженнями та звітами міжнародних організацій. У зв’язку з цим виникає актуальне завдання розробки моделей і методів автоматизованого виявлення та аналізу дезінформації на основі використання підходів штучного інтелекту, які забезпечують можливість оперативної обробки великих обсягів даних і виявлення мереж поширення фейків. Вирішення цього наукового завдання створює умови для зміцнення інформаційної безпеки та стійкості суспільства до ворожих інформаційних впливів.

Обґрунтованість і достовірність наукових результатів, висновків і рекомендацій.

Аналіз змісту дисертаційної роботи дає змогу зробити висновок про належну обґрунтованість наукових положень дослідження. Наукові положення та висновки, представлені у дисертації, обґрунтовано теоретичним аналізом, результатами практичного впровадження на підприємстві та у навчальному процесі. Достовірність отриманих наукових і практичних результатів підтверджується коректним використанням відомих наукових методів у сфері обчислювального штучного інтелекту, а саме – машинного навчання та глибокого навчання, що дають змогу класифікувати і кластеризувати великі набори даних. Результати

досліджень аналізувалися та коригувалися під час апробації на представницьких міжнародних конференціях, а також під час рецензування наукових публікацій.

У якості *основних нових наукових результатів дисертаційної роботи* виступають такі:

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.

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

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.

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

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.

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 забезпечує можливість створення систем виявлення місінформації, дезінформації та малінформації.

Поставлене в дисертаційній роботі наукове завдання виконано повністю.

Повнота викладення результатів досліджень в опублікованих працях.

Дослідження, результати яких викладено в дисертації, виконано на кафедрі математичного моделювання та штучного інтелекту Національного аерокосмічного університету “Харківський авіаційний інститут” в рамках виконання науково-дослідних робіт за проектами “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 представлені у 6 вітчизняних та закордонних виданнях, у тому числі у виданнях, що належать до кuartилів Q1 та Q2 у базі даних Scopus, а також доповідалися та обговорювалися на 4th International Workshop of IT-Professionals on Artificial Intelligence, ProfIT AI 2024 (Кембрідж, США, 2024), V Міжнародній науково-практичній конференції IT-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 IT-професіоналів та аналітиків комп’ютерних систем “ProfIT Conference” (Харків, 2021), III Міжнародній науково-практичній конференції IT-професіоналів та аналітиків комп’ютерних систем “ProfIT Conference” (Харків, 2020).

Опубліковані матеріали повністю відображають зміст дисертації та відповідають вимогам пункту 8 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 р. № 44.

Значимість отриманих результатів для науки і практичного використання. Значимість отриманих результатів для науки підтверджується їх впровадженням у науково-дослідницьку та навчальну діяльність кафедри математичного моделювання та штучного інтелекту Національного аерокосмічного університету “Харківський авіаційний інститут” в рамках виконання науково-дослідних робіт.

Розроблені моделі, методи та програмне забезпечення використані у Balsillie School of International Affairs (акт впровадження від 18 березня 2025 року), Center for International Governance Innovation (акт впровадження від 18 березня 2025 року), Державній установі “Харківський обласний центр контролю та профілактики хвороб при Міністерстві охорони здоров’я” (акт впровадження від 2 грудня 2024 року), а також у освітньому процесі (акт впровадження від 30 січня 2025 року) та науковій діяльності (акт впровадження від 4 березня 2025 року) Національного аерокосмічного університету “Харківський авіаційний інститут”.

Оцінка змісту дисертаційної роботи. Дисертація складається з анотації, змісту, переліку умовних скорочень, вступу, чотирьох розділів, висновку, списку використаних джерел та додатків. Повний обсяг роботи становить 207 сторінок друкованого тексту, з них анотація – на 7 стор., зміст – на 4 стор., перелік умовних скорочень – на 2 стор., основний текст – на 124 стор., список із 151 використаних джерел – на 17 стор., додатки – на 22 стор. Дисертація містить 54 рисунки, та 16 таблиць.

Дисертаційна робота є самостійно виконаною кваліфікаційною науковою працею, яка містить логічно пов’язані наукові та практичні результати, а саме моделі, методи й фреймворк, що реалізує інформаційну технологію виявлення і комплексного аналізу текстової дезінформації та пропаганди в соціальних мережах для ефективного управління стратегічними комунікаціями та протидії ворожим інформаційним атакам. Дисертація виконана з дотриманням вимог академічної доброчесності, отримані результати дають підстави стверджувати про оригінальність роботи.

У вступі обґрунтовано актуальність теми дисертаційної роботи, сформульовано мету, об’єкт, предмет і завдання дослідження, представлено

методи досліджень, наукову новизну отриманих результатів, повноту їх опублікування й апробацію, а також наведено зв'язок дисертаційної роботи з науковими програмами та темами.

Перший розділ дослідження присвячено аналізу інформаційних маніпуляцій у контексті гібридних воєн, зокрема дезінформації, місінформації та малінформації, а також їх ролі у когнітивній та інформаційній війні, а також кібервійні. Розглянуто використання цих форм інформаційних розладів державними та недержавними акторами, зокрема росією, для підриву демократичних процесів через соціальні мережі. Виконано аналіз технологій виявлення та протидії дезінформації з використанням методів машинного і глибокого навчання, трансформерних моделей і гібридних підходів, які мають достатньо високу ефективність у класифікації фейкових новин. Проаналізовано існуючі технологічні фреймворки, спрямовані на моніторинг, класифікацію та візуалізацію поширення маніпулятивного контенту, включно з мультимодальними та причинно-наслідковими підходами. Виконаний аналіз обґрунтовує важливість розробки автоматизованих підходів до протидії інформаційним загрозам.

Другий розділ дисертації присвячено порівняльному аналізу ефективності методів машинного навчання для виявлення текстової дезінформації та пропаганди, зокрема в контексті російських наративів у X (Twitter) та фейкових новин про COVID-19. Розглянуто як класичні моделі (LR, NB, KNN, SVM), так і ансамблеві підходи (BRF, XGBoost, LightGBM), які було протестовано на кількох корпусах, зокрема WELFake і китайськомовному CHECKED з Weibo. Аналіз показав, що найвищу точність продемонстрували моделі SVM і LightGBM, що свідчить про їхню здатність узагальнювати шаблони дезінформації навіть у багатомовних і тематично складних умовах. Також детально описано процеси попередньої обробки даних, балансування класів та оптимізації гіперпараметрів із використанням бібліотеки Optuna. Виконаний аналіз обґрунтовує необхідність детального налаштування моделей, застосування комплексних фреймворків і врахування специфіки мови та платформи для ефективної автоматизованої боротьби з інформаційними загрозами.

У третьому розділі дисертації удосконалено сучасні моделі глибокого навчання для класифікації текстової дезінформації та пропаганди. Розглянуто архітектури BERT, XLNet, LSTM, BiLSTM, Attention-based BiLSTM, а також запропоновано нову гібридну модель, що

поєднує BERT-ембединги з BiLSTM та механізмом уваги. Для кожної моделі виконано експериментальні дослідження на відповідних наборах даних, таких як Fake-Real News, WELFake, CoVID19-FNIR та корпус твітів з російською пропагандою. Порівняння метрик точності, повноти, прецизійності та міри F1 показало, що запропонована модель демонструє найвищу ефективність у задачі класифікації дезінформації завдяки глибокому контекстному розумінню тексту, адаптивній увазі до важливих ознак та оптимізації гіперпараметрів.

У четвертому розділі дисертації представлено розроблений фреймворк для комплексного аналізу текстової дезінформації та пропаганди, який об'єднує класифікацію (на основі моделей BERT, BiLSTM та механізму уваги) з тематичним моделюванням (BERTopic) для виявлення ключових наративів у текстах. Запропоновано нові методи попередньої обробки, ембедингу та кластеризації, що забезпечують високий рівень інтерпретованості результатів. Ефективність фреймворку підтверджено на різних наборах даних: фейкові новини (WELFake), дезінформація про COVID-19 (CoVID19-FNIR), українськомовна дезінформація з Telegram під час повномасштабного вторгнення, а також прокремлівська дезінформація щодо президентських виборів у США. Запропоновані моделі досягли високих значень точності (до 99.47%) та продемонстрували здатність виявляти тонкі семантичні й ідеологічні патерни. Якісний аналіз тематичних кластерів дозволив ідентифікувати ключові наративи (зокрема, змови, делегітимацію еліт, антивакцинні повідомлення), що використовуються у медіакампаніях, та встановити їхню відповідність до стратегічних цілей інформаційного впливу.

У висновках узагальнено отримані в дисертаційній роботі результати, їх новизну та практичну значимість.

Список використаних джерел характеризується достатньою для обраного предмету дослідження кількістю сучасних наукових праць вітчизняних та закордонних науковців і має тісний зв'язок із завданнями, що вирішуються в дисертаційній роботі."

У додатку А надано список публікацій здобувача, у додатку Б - відомості про апробацію результатів, у додатку В - копії актів впровадження наукових результатів, у додатку Г - описи наборів даних використаних для дослідження.

Дотримання вимог академічної доброчесності. Дисертація виконана з дотриманням вимог щодо академічної доброчесності. Отримані результати є оригінальними. У тексті не виявлено використання ідей інших науковців без посилання на їх публікації. Усі наукові та практичні результати опубліковані у необхідному обсязі у закордонних періодичних виданнях, а також апробовані на міжнародних наукових конференціях. Таким чином, порушень академічної доброчесності в дисертаційній роботі та наукових працях Падалко Г.А., де висвітлені основні наукові результати, не виявлено.

Зауваження до дисертаційної роботи.

Зауваження щодо дисертаційної роботи полягають у наступному.

1. У підрозділі 2.1 дисертації детально описані параметри моделі LightGBM, тоді як для SVM/KNN специфікацію параметрів, зокрема тип ядра, значення регуляризації, метрику відстані тощо не наведено.

2. У розділі 3 при описі моделей BERT та XLNet концепції трансферного навчання та механізму уваги відповідно розглядаються як ключові для вирішення задач автоматизованого виявлення і класифікації текстової дезінформації, пропаганди та фейкових новин, але попередній опис цих концепцій в роботі не представлено.

3. У розділі 3 на графіку динаміки точності XLNet (рис. 3.4) спостерігаються суттєві коливання між епохами, зокрема різке зниження точності на 9-й епосі та подальше її відновлення, що може свідчити про відсутність механізмів регуляризації та потребу у подальшому дослідженні впливу регуляризаційних технік на стабілізацію навчання.

4. У розділі 4 (с. 146) авторка зазначає, що було проведено систематичне порівняння дев'яти моделей Sentence-Transformer із вибором моделі All-MiniLM-L6-v2 як оптимальної. Проте у роботі не представлено кількісних метрик (наприклад, точність, F1, час обробки), які б підтверджували переваги цієї моделі над іншими кандидатами.

5. Тематичні кластери, описані в розділі 4, оцінювалися на власних авторських корпусах, без зовнішніх тестових наборів, що ускладнює узагальнення результатів, зокрема при використанні інших мов чи соціальних мереж.

Наведені недоліки не впливають на загальну позитивну оцінку виконаної роботи та не знижують цінність отриманих автором наукових та практичних результатів.

Висновок. Дисертаційна робота Падалко Галини Анатоліївни на тему “Моделі, методи та інформаційна технологія виявлення і аналізу текстової дезінформації та пропаганди в соціальних мережах” за своїм змістом відповідає спеціальності 122 Комп'ютерні науки. Дисертація є завершеною науково-дослідною роботою, яка розв'язує важливе наукове завдання розроблення методів та інформаційної технології, орієнтованих на підсилення стратегічних комунікацій країни на основі виявлення і аналізу текстової дезінформації й пропаганди.

Дисертаційна робота Падалко Галини Анатоліївни на тему “Моделі, методи та інформаційна технологія виявлення і аналізу текстової дезінформації та пропаганди в соціальних мережах” за змістом, структурою, обсягом та оформленням відповідає вимогам Наказу МОН України №40 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 Кабінету Міністрів України від 12.01.2022 №44. Таким чином, здобувач Падалко Галина Анатоліївна заслуговує на присудження наукового ступеню доктора філософії у галузі знань 12 Інформаційні технології за спеціальністю 122 Комп'ютерні науки.

Офіційний опонент:

Доктор технічних наук,
професор, професор кафедри
інформаційних управляючих систем
Харківського національного
університету радіоелектроніки

Сергій ЧАЛИЙ