

## 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 дисертації**

У контексті триваючої гібридної агресії, яку Росія веде проти України та демократичних країн Заходу, дезінформація й пропагандистські наративи набули статусу основних засобів інформаційного тиску. Особливу роль у цьому відіграють соціальні медіа, які забезпечують швидке розповсюдження маніпулятивних матеріалів та ускладнюють виявлення джерел інформаційних атак. Кремлівський режим систематично використовує інформаційні кампанії з метою підриву легітимності державних інституцій, загострення внутрішніх протиріч і послаблення довіри громадян до офіційних джерел. Ці дії ретельно документуються у звітах міжнародних організацій та дослідницьких центрів. У зв’язку з цим постає нагальна потреба у впровадженні автоматизованих інструментів на основі штучного інтелекту, здатних ефективно обробляти великі масиви даних і виявляти мережі поширення фейкової інформації. Вибір теми зумовлений актуальністю завдань зміцнення інформаційної безпеки та посилення стійкості суспільства до зовнішнього інформаційного тиску.

### **Оцінка обґрунтованості наукових результатів дисертації, їх достовірності та новизни**

- 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, 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.

- 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, що 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 дозволяє створення систем для виявлення місінформації, дезінформації та малінформації.

Наукова достовірність результатів дисертаційного дослідження, а також обґрунтованість сформульованих висновків і рекомендацій, базується на використанні сучасних концепцій обчислювального штучного інтелекту, зокрема алгоритмів машинного й глибокого навчання, що дозволяють ефективно аналізувати, класифікувати та групувати великі масиви даних. Отримані результати проходили апробацію в рамках престижних міжнародних наукових конференцій та були піддані рецензуванню у спеціалізованих наукових виданнях, що підтверджує їхню наукову вагомість і практичну значущість. Отже, поставлене в дисертаційній роботі наукове завдання виконано повністю, здобувачка повною мірою оволоділа методологією наукової діяльності.

## **Оцінка змісту дисертації, її завершеність та дотримання принципів академічної доброчесності**

За своїм змістом дисертаційна робота здобувачки Падалко Г.А. повністю відповідає стандарту вищої освіти зі спеціальності 122 Комп'ютерні технології та напрямкам досліджень відповідно до освітньої програми «Інформаційні технології». Дисертаційна робота є завершеною науковою працею і свідчить про наявність особистого внеску здобувачки у науковий напрям.

Розглянувши звіт подібності за результатами перевірки дисертаційної роботи на текстові співпадиння, можна зробити висновок, що дисертаційна робота Падалко Галини Анатоліївни є результатом самостійних досліджень здобувачки і не містить елементів фальсифікації, компіляцій, фабрикації, плагіату та запозичень. Використані ідеї, результати і тексти інших авторів мають належні посилання на відповідне джерело. Виявлений відсоток співпадинь пояснюється посиланнями на уже опубліковані роботи здобувачки. Дисертація написана з дотриманням принципів академічної доброчесності.

## **Мова та стиль викладення результатів**

Дисертаційна робота написана українською мовою. Текст дисертації викладений логічно і послідовно, його оформлення відповідає чинним вимогам. Автор дотримується наукового стилю та використовує загальноприйняту термінологію.

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.

У вступі дисертаційної роботи висвітлено актуальність обраної тематики, чітко окреслено мету, об'єкт, предмет і основні завдання дослідження. Також подано огляд використаних методологічних підходів, визначено наукову новизну результатів, ступінь їх апробації та оприлюднення, а також вказано на зв'язок дисертаційного дослідження з чинними науковими програмами та проєктами.

Перший розділ присвячено ґрунтовному теоретичному аналізу феномену інформаційних маніпуляцій у контексті гібридних війн. Розглянуто природу та відмінності між дезінформацією, місінформацією й малінформацією, а також їхню роль у когнітивних, інформаційних та кіберконфліктах. Особливу увагу приділено тому, як державні й недержавні суб'єкти — зокрема росія — використовують ці інструменти через соціальні мережі, штучний інтелект і стратегічні кампанії для підриву демократичних процесів. Окремо проаналізовано сучасні технології виявлення та протидії дезінформації: методи машинного і глибокого навчання, трансформерні архітектури, гібридні підходи. Також розглянуто наявні технологічні фреймворки для моніторингу, класифікації та візуалізації поширення шкідливого контенту, включаючи мультимодальні й каузальні підходи. Розділ формує концептуальне та технічне підґрунтя для подальшої розробки автоматизованих систем протидії інформаційним загрозам.

Другий розділ зосереджено на порівнянні ефективності різних методів машинного навчання для виявлення текстової дезінформації та пропаганди. Проаналізовано класичні алгоритми (логістична регресія, наївний баєсівський класифікатор, KNN, SVM) і сучасні ансамблеві моделі (BRF, XGBoost, LightGBM), які були протестовані на низці корпусів, зокрема WELFake та CHECKED (з китайською мовою платформи Weibo). Моделі SVM та LightGBM продемонстрували найвищі показники точності — 82% для Twitter, 89% для Weibo, а також F1-міру 0.977 для WELFake. У роботі докладно описано підходи до попередньої обробки даних, балансування класів і гіперпараметричної оптимізації з використанням фреймворку Optuna. Отримані результати підтверджують важливість адаптації моделей до мовної та платформної специфіки.

Третій розділ присвячено аналізу й удосконаленню архітектур глибокого навчання для задач класифікації дезінформаційного контенту. Розглянуто моделі BERT, XLNet, LSTM, BiLSTM та Attention-based BiLSTM, а також запропоновано нову гібридну модель, що інтегрує BERT-ембединги з BiLSTM і механізмом уваги. Моделі були протестовані на різних датасетах, включаючи Fake-Real News, WELFake, CoVID19-FNIR та корпус твітів із прокремлівською пропагандою. Порівняльний аналіз за метриками точності, повноти, прецизійності та F1-міри засвідчив перевагу запропонованої моделі, яка

досягла точності 99.47%. Це стало можливим завдяки контекстному розумінню, адаптивній увазі до ключових ознак та ефективній оптимізації.

У четвертому розділі представлено інтегрований фреймворк для поглибленого аналізу дезінформаційного тексту, який поєднує класифікаційні моделі (BERT, BiLSTM з attention-механізмом) із тематичним моделюванням (BERTopic) для виявлення ключових ідеологічних наративів. Запропоновано покращені підходи до препроцесингу, векторизації та кластеризації, що підвищують інтерпретованість результатів. Експерименти на корпусах WELFake, CoVID19-FNIR, українськомовному Telegram-контенті періоду повномасштабного вторгнення та прокремлівських повідомленнях про вибори в США підтвердили високу точність (до 99.47%) і здатність моделі розпізнавати приховані семантичні й ідеологічні патерни. Тематичний аналіз дозволив ідентифікувати ключові наративи — теорії змови, делегітимацію еліт, антивакцинні меседжі — й встановити їхню відповідність стратегічним цілям впливу.

У висновках підсумовано основні результати дисертаційної роботи, наголошено на їхній науковій та практичній цінності, а також запропоновано напрями подальших досліджень у сфері автоматизованої боротьби з дезінформацією.

Список використаних джерел охоплює актуальні публікації провідних вітчизняних і міжнародних дослідників, які безпосередньо пов'язані з темою дисертаційного дослідження.

У додатках представлено:

- додаток А — перелік публікацій здобувачки;
- додаток Б — інформацію про апробацію результатів;
- додаток В — копії актів впровадження;
- додаток Г — опис використаних наборів даних.

#### **Недоліки та зауваження до дисертаційної роботи**

1. Опис пайплайну BERTopic подано на рівні концептуальної схеми, без покрокового алгоритму чи псевдокоду, що ускладнює точне відтворення.
2. Робота декларує, що фреймворк «гнучкий і модульний», проте не містить чек-листа дій (Data → Model → KPI).
3. У підрозділах 4.1 – 4.3 описано, що ембединги BERT зменшуються UMAP-ом, кластеризуються HDBSCAN та допрацьовуються KeyBERT,

але процедура подана лише описово, без покрокового алгоритму чи псевдокоду

4. Перелік скорочень містить невизначене MDM
5. В роботі присутні граматичні та стилістичні помилки.

Вважаю, що висловлені зауваження не є визначальними і не зменшують загальну наукову новизну та практичну значимість результатів та не впливають на позитивну оцінку дисертаційної роботи.

### **Загальні висновки по дисертаційній роботі**

Дисертаційна робота Падалко Галини Анатоліївни на тему «Моделі, методи та інформаційна технологія виявлення і аналізу текстової дезінформації та пропаганди в соціальних мережах», подана на здобуття ступеня доктора філософії за спеціальністю 122 Комп'ютерні науки 12 Інформаційні технології є самостійною та завершеною працею, містить елементи наукової новизни та практичну цінність.

Дисертаційна робота Падалко Галини Анатоліївни 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 КМУ від 12.01.2022 №44.

Таким чином, Падалко Галина Анатоліївна, авторка дисертаційної роботи на тему «Моделі, методи та інформаційна технологія виявлення і аналізу текстової дезінформації та пропаганди в соціальних мережах», заслуговує на присудження наукового ступеня доктора філософії за спеціальністю 122 Комп'ютерні науки.

### **Офіційний опонент:**

кандидат технічних наук,  
доцент, доцент кафедри соціальних  
комунікацій та інформаційної діяльності  
Національного університету  
“Львівська політехніка”

Юрій ССРОВ