Голові разової спеціалізованої
вченої ради PhD 8733
Хмельницького національного університету
доктору технічних наук, професору
Тетяні ГОВОРУЩЕНКО

РЕЦЕНЗІЯ
на дисертаційне дослідження Собко Олени Віталіївни
на тему «Методи виявлення та класифікації кіберзалякувань у текстовому контенті засобами штучного інтелекту»,
подане на здобуття ступеня доктора філософії
з галузі знань 12 Інформаційні технології
за спеціальністю 122 Комп'ютерні науки

1. Актуальність теми та зв'язок з науковими планами наукових робіт університету.

Актуальність тематики дисертаційного дослідження зумовлена стрімким зростанням кількості випадків кіберзалякувань у цифровому середовищі, особливо в контексті широкого використання соціальних мереж, платформ онлайн-комунікації та цифрової освіти. Проблематика виявлення й класифікації такого типу агресивної поведінки має не лише наукове, а й безпосереднє прикладне значення в рамках розробки інтелектуальних систем захисту та моніторингу контенту.

Дисертаційна робота орієнтована на розв'язання задач, пов'язаних із аналізом україномовного текстового контенту з використанням сучасних моделей глибокого навчання, зокрема BERT та BiLSTM, а також засобів пояснювального штучного інтелекту. Застосування принципів FATE у побудові інтелектуальних систем класифікації контенту демонструє підхід, що охоплює технічні, етичні та соціальні аспекти, і є надзвичайно актуальним у світовому дослідницькому просторі.

Дослідження виконано в межах науково-дослідної роботи Хмельницького національного університету за держбюджетною темою «Розроблення інформаційної технології прийняття контрольованих людиною критично-безпекових рішень за ментально-формальними моделями машинного навчання» (ДР № 0121U112025), у якій автор брала безпосередню участь, розробляючи архітектури нейромережевих моделей та інструментарій їхньої інтерпретації. Таким чином, тематика дисертаційної роботи безпосередньо пов'язана з

пріоритетними напрямами наукових досліджень університету у сфері штучного інтелекту, інформаційної безпеки та мовних технологій, а результати можуть бути впроваджені як у навчальний процес, так і в прикладні системи протидії кіберзалякуванням.

2. Формулювання наукової задачі, мети й задач дослідження.

Авторкою правильно визначено об'єкт і предмет дослідження, відповідно до сформульованої науко-прикладної задачі. Так, об'єктом дослідження визначено процес інтелектуального аналізу текстового контенту для виявлення кіберзалякувань. Предметом дослідження є методи та засоби обробки природної мови для виявлення кіберзалякувань у текстовому контенті. Сформульовано науково-прикладну задачу забезпечення автоматизованих виявлення та класифікації кіберзалякувань з подальшою інтерпретацією прийнятих рішень, шляхом до розв'язання якої пропонується розробка методів і засобів виявлення та класифікації кіберзалякувань у текстовому контенті, що сприятиме підвищенню точності та якості виявлення кіберзалякувань у текстовому контенті з подальшою інтерпретацією прийнятих рішень.

Мету дисертаційної роботи визначено, як підвищення точності та якості виявлення кіберзалякувань у текстовому контенті засобами штучного інтелекту з подальшою інтерпретацією прийнятих рішень.

Поставлену мету роботи досягнуто в результаті розв'язання таких задач: 1) проведено аналіз методів, засобів та технологій для автоматизованого виявлення кіберзалякувань у текстовому контенті; 2) розроблено новий метод оцінювання та коригування репрезентативності датасету за FATE-принципом справедливості, що забезпечуватиме недискримінацію за віковою, гендерною і релігійною приналежністю; 3) розроблено новий метод виявлення кіберзалякувань у текстовому контенті; 4) удосконалено метод інтерпретації результатів виявлення кіберзалякувань; 5) створено інтелектуальну інформаційну систему для валідації розроблених методів і провести експерименти та порівняння.

3. Наукова новизна одержаних авторкою результатів полягає в наступному:

1) Вперше запропоновано метод оцінювання та коригування репрезентативності датасету за FATE-принципом справедливості, що забезпечує недискримінацію за віковою, гендерною, релігійною приналежністю, що дозволило підвищити якість навчання класифікаторів для виявлення кіберзалякувань.

2) Розроблено новий метод виявлення кіберзалякувань у текстовому контенті, який відрізняється від існуючих двоетапним виявленням кіберзалякувань, що полягає у нейромережевій ідентифікації наявності кіберзалякувань та подальшій нейромережевій мультилейбловій класифікації окремих типів кіберзалякувань, що дало можливість підвищити точність та якість виявлення кіберзалякувань.

3) Удосконалено метод інтерпретації результатів виявлення кіберзалякувань, який відрізняється від існуючих можливістю надавати візуальні пояснення для мультилейблової класифікації виявлених типів кіберзалякувань в альтернативних поданнях.

4. Обґрунтованість і достовірність наукових положень, висновків і рекомендацій.

Наукові положення, висновки й рекомендації дисертації обґрунтовані коректним та доцільним використанням сучасного математичного апарату, методів глибокого навчання, інструментів пояснювального штучного інтелекту, а також результатами реалізації прототипу інтелектуальної інформаційної системи для виявлення кіберзалякувань в україномовному текстовому контенті. Експериментальні дослідження, проведені на репрезентативних текстових корпусах, підтвердили ефективність і стабільність роботи розробленої системи, а впровадженням результатів дисертаційної роботи на підприємствах засвідчило узгодженість отриманих теоретичних результатів із реальними практичними потребами.

5. Практичне значення одержаних результатів.

Практичне значення результатів дисертаційної роботи полягає у розробленні та успішній реалізації інтелектуальної інформаційної системи виявлення та класифікації кіберзалякувань у текстовому контенті засобами штучного інтелекту. Запропонована система базується на інтеграції авторських методів оцінювання та коригування репрезентативності навчальних вибірок відповідно до етичних вимог FATE-принципу справедливості, механізмів виявлення та мультилейблової класифікації кіберзалякувань, а також засобів інтерпретації отриманих результатів. Розроблена система забезпечує не лише підвищення точності й якості автоматизованого виявлення агресивного контенту, але й формує прозорі, візуально пояснені рішення, що істотно підвищує довіру до результатів класифікації. Отримані наукові результати мають прикладну цінність і можуть бути використані у проєктуванні сучасних систем інформаційної безпеки, контент-моніторингу та освітніх платформ, що функціонують у цифровому середовищі.

Результати дисертаційної роботи впроваджено: у відділі протидії кіберзлочинам у Хмельницькій області Департаменту кіберполіції Національної поліції України; у ПП «Авіві» (довідка про впровадження); у ГО «ІТ-кластер міста Хмельницького» (довідка про впровадження); у ТОВ «Системи для бізнесу 2» (довідка про впровадження); у навчальному процесі Хмельницького національного університету (акт впровадження); при виконанні держбюджетної теми Хмельницького національного університету «Розроблення інформаційної технології прийняття контрольованих людиною критично-безпекових рішень за ментально-формальними моделями машинного навчання» (ДР № 0121U112025)

6. Особистий внесок здобувачки полягає у наступному:

Розроблені методи та засоби забезпечують розв'язання поставлених у дисертації задач. Усі основні наукові та прикладні результати дисертаційної роботи отримані здобувачкою самостійно.

Основні наукові результати, отримані у дисертаційній роботі, пройшли належну апробацію та оприлюднення, що свідчить про їх відповідність сучасним вимогам до наукових досліджень. Зокрема, опубліковано 11 наукових праць, серед яких 4 статті у фахових наукових журналах України, що входили на момент публікації до переліку наукових фахових видань України категорії Б, 5 публікацій у виданнях, що індексуються у міжнародній наукометричній базі даних Scopus (засвідчують апробацію матеріалів дисертації), а також отримано 2 авторські свідоцтва на результати інтелектуальної діяльності. Такий рівень публічної апробації підтверджує достатню наукову новизну, практичну значущість та достовірність результатів, викладених у дисертації. У роботах, що опубліковані в співавторстві, авторці належать основні ідеї, теоретична та практична розробка положень, які відображені у характеристиці наукової новизни отриманих результатів, а саме: метод оцінювання та коригування репрезентативності датасету за FATE-принципом справедливості; метод виявлення кіберзалякувань у текстовому контенті; метод інтерпретації результатів виявлення кіберзалякувань.

7. Апробація матеріалів дисертації.

Основні результати дисертаційного дослідження доповідались та обговорювались на міжнародних науково-практичних конференціях та семінарах: 6th International Conference on Computational Linguistics and Intelligent Systems «CoLInS 2022» (12–13 May, 2022, Gliwice, Poland); 16th International Scientific Conference «Intellectual Systems of Decision-Making and Problems of Computational Intelligence ISDMCI-2022)» (June 14–16, 2022, Rivne, Ukraine); Intelligent Systems Workshop at 8th International Conference on

Computational Linguistics and Intelligent Systems «ISW-CoLInS 2024» (April 12–13, 2024, Lviv, Ukraine); 1st International Workshop at Advanced Applied Information Technologies «AdvAIT 2024» (December 5, 2024, Khmelnytskyi, Ukraine, Zilina, Slovakia); 7th Workshop for Young Scientists in Computer Science & Software Engineering «CS&SE@SW 2024» (December 27, 2024, Kryvyi Rih, Ukraine).

8. Структура та обсяг дисертації.

Дисертаційна робота складається з анотації, змісту, переліку умовних скорочень, вступу, чотирьох розділів, висновків, списку використаних джерел із 162 найменувань на 22 сторінках і 4 додатків. Загальний обсяг дисертаційної роботи становить 174 сторінки друкованого тексту, із них 137 сторінок основного тексту. Дисертація містить 45 рисунків та 11 таблиць.

9. Зауваження.

У результаті вивчення рукопису мною сформовано такі зауваження:

1. У вступі (с. 11) зазначено факт упровадження результатів дисертаційної роботи у навчальний процес Хмельницького національного університету, проте не конкретизовано форми та механізми такого впровадження, зокрема не наведено інформації щодо дисциплін, навчальних програм чи програмного забезпечення, в яких були використані результати дослідження. Варто деталізувати цей аспект для повнішої характеристики практичного значення роботи.
2. Розроблений у роботі метод оцінювання та коригування репрезентативності датасету за FATE-принципом справедливості має власну наукову цінність та призначення, проте в назві дисертації ніяким чином не відображено ролі цього методу.
3. Надмірна декларативність без доказової бази (с. 3, останній абзац): твердження про те, що модель «підвищує якість» або «забезпечує недискримінацію» не супроводжуються статистичними підтвердженнями в анотації.
4. Відсутність методологічної обґрунтованості щодо вибору FATE саме як базового принципу етики (п.п. 1.1, 1.2, 2.1): не надано аргументів, чому інші класифікації (наприклад, FATML) не були розглянуті.
5. Не вказано, як забезпечується стійкість моделі до сарказму, що є частою проблемою в кіберзалякуваннях: чи розглядаються метафори, багатозначність?
6. Необґрунтованість обрання моделі LIME для інтерпретації (п. 1.4): чому не SHAP, який дає більш точну оцінку внеску ознак?

Проте підкреслюю, що зазначені зауваження істотно не впливають на зміст дисертаційної роботи та не знижують її наукову новизну та практичну цінність.

10. Загальний висновок.

Дисертаційна робота Собко О.В. «Методи виявлення та класифікації кіберзалякувань у текстовому контенті засобами штучного інтелекту» є завершеною науковою роботою, яка містить новий та актуальний науково-прикладний внесок. Усі результати, які виносяться на захист, є достовірними та отримані авторкою особисто.

Тому, з огляду на вище вказане, вважаю, що дисертаційна робота «Методи виявлення та класифікації кіберзалякувань у текстовому контенті засобами штучного інтелекту», яка подана на здобуття ступеня доктора філософії, за своїм науковим рівнем та практичною цінністю, змістом та оформленням повністю відповідає вимогам пп. 6, 7, 8, 9 «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 січня 2022 р. №44 (зі змінами, внесеними згідно з Постановами Кабінету Міністрів України № 341 від 21.03.2022, № 502 від 19.05.2023, № 507 від 03.05.2024), а її авторка, Собко Олена Віталіївна, заслуговує на присудження ступеня доктора філософії за спеціальністю 122 Комп'ютерні науки.

Рецензент
доктор технічних наук, професор
кафедри комп'ютерної інженерії
та інформаційних систем
Хмельницького національного університету _____ Сергій ЛИСЕНКО

«Підпис Сергія ЛИСЕНКО засвідчую»:
Проректор з наукової роботи
Хмельницького національного університету _____ Олег СИНЮК