РЕЦЕНЗІЯ

на дисертаційну роботу **Панчука Богдана Олександровича** **«Виявлення мережевих атак алгоритмами штучного інтелекту»,** подану на здобуття наукового ступеня доктора філософії за спеціальністю **122 Комп'ютерні науки.**

1. Актуальність теми дослідження.

Інформаційні системи є інтегрованими майже у всі сфери функціонування підприємств та держави, а їх безпека стає першочерговим завданням для дослідників та інженерів. Однією з найбільш актуальних задач сьогодення є розробка та вдосконалення засобів протидії шкідливим програмам та засобам, які становлять значну загрозу для окремих громадян і для країни в цілому. Експерти в галузі кібербезпеки постійно стикаються з проблемою появи нових, досконаліших типів вірусних програм, здатних самостійно розповсюджуватись мережею та маскувати свою активність. Це ускладнює їх виявлення методами, що ґрунтуються на наборах правил співставлення, сформованих на основі попередньої експертизи та аналізу шаблонів поведінки вже відомих прикладів шкідливих програм. Альтернативою до цього підходу є використання методів штучного інтелекту, які здатні самостійно виявляти неочевидні закономірності у мережевих даних. Такі властивості роблять їх перспективним інструментом для побудови адаптивних та більш стійких систем кіберзахисту, що обумовлює актуальність запропонованої здобувачем теми дослідження.

2.Зв'язок дослідження з наукові програмами та теми.

Дисертаційна робота виконана у відповідності з планами науково-дослідних робіт Інституту кібернетики імені В.М. Глушкова, а саме при виконанні НДР ВП.100.17 «Розробити формальні методи виявлення зловмисної поведінки в мережі та хмарному оточенні на основі комбінації алгебраїчних методів та машинного навчання» (державний реєстраційний номер 0122U001164).

3. Наукова новизна одержаних результатів.

У дисертаційній роботі розглянуто малодосліджену задачу стійкості систем виявлення мережевих загроз на основі ШІ до можливих змагальних впливів з боку зловмисника. Запропонована нова адаптація швидкого методу знаку градієнту для генерації змагальних прикладів шкідливих мережевих даних, які можуть спричиняти помилки класифікації. Цей метод використовується для оцінки стійкості нейромережі до можливих збурень. Було продемонстровано можливість доповнення навчальної вибірки штучно

згенерованими даними з метою підвищення показника стійкості моделі до змагальних атак.

Особливу увагу приділено розробці методів формальної верифікації нейронних мереж для отримання більш достовірних гарантій якості систем кібербезпеки. В роботі вперше формалізується критерій локальної стійкості нейромереж-класифікаторів потоків мережевого трафіку. Для верифікації заданого критерію розроблено метод автоматизованого доведення властивостей нейронних мереж із використанням SMT-розв'язувача. Метод ґрунтується на новому алгоритмі спрощення символьного представлення обчислювального графа нейромережі шляхом формального доведення можливості тотожної заміни кусково-лінійних функцій активації лінійними функціями.

4. Практичне значення отриманих результатів

У роботі розв'язується актуальна прикладна задача розробки та підвищення ефективності систем виявлення трафіку шкідливого ПЗ на основі аналізу потоків мережевих даних на основі моделей штучного інтелекту. Створено розширену навчальну вибірку з прикладами трафіку різноманітних вірусних програм та продемонстровано ефективність їх виявлення класифікаторами нейронних мереж та алгоритмами машинного навчання. Для більш достовірної оцінки якості системи в роботі було сформовано нову тестову вибірку, шляхом комбінацій кількох відкритих наборів трафіку. Показано, що при використанні нейронних мереж та алгоритму випадкового лісу, повнота виявлення шкідливого трафіку сягає 80%, що свідчить про практичну цінність розробленої системи.

5. Ступінь обґрунтованості основних положень та висновків дисертації.

Основні наукові положення, висновки, рекомендації дисертаційної роботи мають належне теоретичне, методологічне та емпіричне обґрунтування. Матеріали досліджень викладені у роботі послідовно та зрозуміло.

Теоретичні положення, що представлені у роботі, мають експериментальне підтвердження. Експерименти поставлені коректно, а їх результати подані структуровано та зрозуміло у таблицях та графіках.

6. Повнота викладення наукових положень та висновків в опублікованих працях.

Сутність основних отриманих результатів виконаного дослідження та їх наукова новизна достатньо повно відображені у п'яти наукових статтях, дві з яких входять до наукометричної бази даних Scopus.

Публікації відповідають вимогам до наукових статей, встановлених МОН України. Результати проведеного дослідження доповідалися на міжнародній науковій конференції.

7. Недоліки дисертації щодо її змісту і оформлення.

Суттєвих недоліків щодо змісту і оформлення дисертації не вбачаю. Наведу лише декілька зауважень.

1). У розділі 3.2 для знаходження збурень, до яких вразлива модель, при генерації змагальних прикладів робиться лише один крок в напрямку градієнту функції втрат нейромережі. Такий підхід є ефективним лише для невеликих значень ε. Для пошуку більш оптимальних прикладів варто було б використовувати багатокрокові (ітеративні) методи з перерахунком значень градієнту на кожному новому кроці.

2). У характеристичному векторі мережевих потоків протоколу UDP присутні надлишкові атрибути специфічні для протоколу TCP. Наприклад, «кількість SYN прапорців», «розмір TCP вікна» та інші. Це може негативно впливати на ефективність класифікації UDP трафіку. Для кожного набору характеристичних ознак варто було б провести навчання окремої моделі і під час класифікації динамічно обирати відповідну в залежності від типу транспортного протоколу.

3). У розділі 4.6 варто було б навести результати формальної верифікації моделі класифікації навченої не лише на оригінальному наборі даних, а і на доповненому, створеному в розділі 3. Це дозволило б показати підвищення стійкості нової моделі до можливих збурень у вхідних даних.

Наведені зауваження в жодному разі не впливають на загальну позитивну оцінку дисертаційної роботи.

9.Рекомендації щодо впровадження результатів дисертаційного дослідження в практику.

Отримані методи і алгоритми можуть бути застосовані в інформаційних системах різного призначення при аналізі потоків мережевих даних для виявлення мережевих загроз.

Результати дослідження можуть бути включені до навчально-методичного забезпечення відповідних навчальних курсів студентів IT спеціальностей у вигляді лекційних матеріалів, методичних рекомендацій, тощо.

10. Відповідність дисертації встановленим вимогам.

Методи та алгоритми запропоновані у дисертації Панчука Б.О. мають наукову новизну та практичну цінність в області мережевої кібербезпеки. Здобувач робить власний внесок у вирішення проблеми формальної верифікації та підвищення стійкості нейронних мереж до змагальних впливів,

а також експериментально підтверджує ефективність запропонованих підходів і методів задачі класифікації мережевих даних та виявлення трафіку шкідливих програм.

Зважаючи на вищесказане, дисертація **Панчука Богдана Олександровича «Виявлення мережевих загроз алгоритмами штучного інтелекту»** за актуальністю, об'ємом і рівнем проведених досліджень, науковою новизною і практичною значимістю відповідає вимогам «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 №44 від 12 січня 2022 року, а її автор заслуговує присудження ступеня доктора філософії за спеціальністю **122 Комп'ютерні науки**.

Завідувач відділу методів системного моделювання
Інституту кібернетики НАН України
член-кореспондент НАН України,
доктор фізико-математичних наук Володимир ПЕПЕЛЯЄВ

Підпис В. Пепеляєва
ЗАСВІДЧУЮ
Зав. канц.
НАН України 04.06.2025