

## ВІДГУК

офіційного опонента про дисертаційну роботу

ДОЛГИХ Сергія Миколайовича

«Інформаційна технологія розпізнавання мережевих даних Інтернет  
на основі генеративних нейромережевих моделей»,

представлену на здобуття наукового ступеня кандидата технічних наук  
за спеціальністю 05.13. 06 «Інформаційні технології»

**Актуальність.** На сьогоднішній день задачі застосування штучного інтелекту, а саме прийняття складних рішень автоматичними системами, стають ще більш актуальними в різних галузях інформаційних технологій та суспільстві. В першу чергу, це пов'язано з ростом об'ємів інформації, що обробляється, а також таких факторів, які впливають на прийняття рішень фахівцем, як точність, стабільність, обмеження часу і т.д. У багатьох випадках, успіх застосування традиційних систем машинного інтелекту залежить від наявності наборів навчальних даних, які повинні задовольняти деяким суттєвим умовам, а саме:

- навчальні дані пов'язані з відомим результатом, який характеризує проблему, а саме заздалегідь відомі класи, категорії;
- високий рівень надійності в асоціації вхідних даних з відомим результатом;
- представництво, тобто високий рівень відповідності даних набору розподілу реальних даних;
- достатній розмір навчальних наборів та інші.

Проте при аналізі та класифікації даних у мережах Інтернет існують серйозні перешкоди для виконання цих умов і, як наслідок, успішного навчання та застосування загальноприйнятих методів машинного інтелекту. Наприклад, як

було зазначено у кількох опублікованих здобувачем результатах, характеристичні параметри мережного трафіку мають тенденцію до зміни як у часі, так і в залежності від мережі-джерела, і з цієї причини використання доступних наборів може не забезпечити бажаного рівня точності машинних систем у задачах розпізнавання та класифікації звичайними методами. Це спонукає до пошуку методів навчання машинних систем, які меншою мірою залежать від підготовлених значних наборів навчальних даних і здатні використовувати систему характеристичних параметрів і концептів, які відповідають конкретним наборам даних та середовищам використання.

Отже, науково-прикладною задачею дисертаційної роботи є обґрунтування, розробка та експериментальна перевірка методів навчання систем машинного інтелекту, які здатні навчатися з мінімальними наборами даних Інтернет, при збереженні або покращенні характеристик успіху навчання, включаючи точність, стабільність та інші.

Зважаючи на вищевикладене, вважаю, що тема дисертаційного дослідження Долгих С.М. є актуальною.

**Структура та обсяг дисертації.** Дисертаційна робота складається з анотації, змісту, переліку умовних скорочень, вступу, чотирьох розділів, загальних висновків, додатків, списку використаних джерел до кожного розділу та має 129 сторінок основного тексту, 29 рисунків, 20 таблиць, 2 сторінки додатків. Загальний обсяг роботи – 139 сторінок.

**Обґрунтованість і достовірність наукових положень, висновків і рекомендацій,** сформульованих у дисертації, підтверджується коректною постановкою завдань, науковою обґрунтованістю теоретичних положень, використанням апробованого математичного апарату, узгодженістю теоретичних положень з результатами експериментальних досліджень, опублікованими

науковими публікаціями у фахових виданнях та відповідним актом впровадження у науково-дослідну діяльність науково-дослідної лабораторії протидії кіберзагрозам в авіаційній галузі, у рамках виконання Науково-дослідної роботи «Інтелектуалізована система захищеного передавання пакетних даних на базі розвідувально-пошукового безпілотного літального апарату».

**Наукова новизна** одержаних особисто здобувачем результатів полягає у наступному:

1. Удосконалено концептуальну і математичну модель розподілів даних пакетів трафіку Інтернет, яка за рахунок використання неконтрольованих генеративних представлень, дозволяє підвищити ефективність виділення інформативних факторів даних.

2. Отримали подальший розвиток методи теорії генеративних представлень у напрямку розробки методів аналізу структури щільності розподілів даних пакетів трафіку Інтернет в представленнях генеративних моделей глибокого навчання, що за рахунок застосування оригінальної архітектури автоенкодера з різким стисненням розмірності шару кодування, забезпечило підвищення ефективності навчання за характеристиками зниження помилки та відтворення даних при неконтрольованому генеративному навчанні з використанням даних Інтернет та інших типів даних.

3. Вперше доведена теорема про категоризацію генеративних представлень, що на підставі методів варіаційного аналізу забезпечує теоретичне обґрунтування методів навчання на основі генеративної структури представлень.

4. Вперше розроблено методи навчання призначені для розпізнавання відомих класів даних пакетів трафіку Інтернет, які за рахунок розроблених концептуальної та математичної моделі, методів теорії генеративних представлень та теореми про категоризацію генеративних представлень,

забезпечують: достатню точність розпізнавання; зменшення залежності від джерела отримання навчальних даних; зменшення обсягу навчальних даних в порівнянні з відомими методами контрольованого навчання.

5. Вперше розроблено методи навчання призначені для розпізнавання натуральних концептів даних пакетів трафіку Інтернет, які за рахунок визначення структури щільності генеративних представлень, дозволяють реалізовувати розпізнавання натуральних концептів даних Інтернет без використання маркованих даних, з точністю на рівні відомих методів контрольованого навчання.

6. Вперше визначено, формалізовано та виконано прототипну реалізацію інформаційної технології розпізнавання класів даних трафіку Інтернет, що за рахунок застосування запропонованих методів визначення структури щільності генеративних представлень, дозволяє автоматизувати процес навчання та використання запропонованих моделей і методів розпізнавання класів трафіку Інтернет.

**Теоретичне та практичне значення роботи.** Запропоновані моделі і методи є науково-методологічною основою для розробки інформаційної технології обробки даних та навчання з наборами даних пакетів трафіку Інтернет мінімального обсягу на основі визначення структури щільності генеративних представлень.

- Проведено порівняльний аналіз сучасних методів розпізнавання та класифікації даних пакетів трафіку Інтернет. Отримано порівняльні результати ефективності та недоліків розглянутих методів.
- Отримано програмну реалізацію моделей генеративного навчання автоенкодера з різким стиском розмірності кодуючого шару кодування у

програмному середовищі Python, що дозволило отримати інформативні представлення даних Інтернет суттєво зменшеної розмірності.

- Отримано програмне втілення масивів даних пакетів трафіка Інтернет та зображень у програмному середовищі Python.
- Отримано програмну реалізацію методів визначення структури щільності генеративних представлень у програмному середовищі Python (методи кластеризації за щільністю, багатовимірних гістограм), що забезпечило стабільне визначення структури щільності генеративних представлень даних Інтернет, з успішністю генеративного навчання та визначення структури щільності вище 80%.
- Отримано програмну реалізацію методів розпізнавання відомих класів даних пакетів трафіку Інтернет з використанням структури щільності генеративних представлень у програмному середовищі Python з використанням пакетів та бібліотек машинного навчання. Дозволяє стабільне навчання розпізнавання відомих класів даних Інтернет при зменшенні обсягу навчальних даних, у 10–100 разів в порівнянні з відомими методами контрольованого навчання.
- Отримано програмну реалізацію методу навчання розпізнавання натуральних концептів даних пакетів трафіку Інтернет з використанням структури щільності генеративних представлень, у програмному середовищі Python з використанням пакетів та бібліотек машинного навчання та обробки даних. Дозволяє стабільне навчання розпізнавання натуральних типів (концептів) даних Інтернету без вимоги відомих даних навчання.
- Отримано програмне прототипне виконання інформаційної технології розпізнавання даних пакетів трафіку Інтернет при наборах навчання мінімального обсягу з використанням структури щільності генеративних

представлень у програмному середовищі Python з використанням пакетів та бібліотек машинного навчання.

**Оцінка змісту дисертації.** В першому розділі були розглянуті та проаналізовані сучасні методи та моделі неконтрольованого навчання. Зроблено висновок, що використання саме генеративних нейромережових архітектур може бути ефективним підходом у створенні інформативних представлень даних та дозволить розв'язати поставлене завдання успішного навчання штучних систем інтелекту з мінімальними відомими даними.

У другому розділі були проаналізовані методи та моделі неконтрольованого генеративного навчання, створено інформативні низькорозмірні представлення даних пакетів трафіку Інтернет та визначено інформативну структуру представлень, таким чином обґрунтовано використання інформативної структури генеративних представлень для розробки методів навчання зі значно зменшеними вимогами до наявності заздалегідь відомих даних навчання.

У третьому розділі наведений детальний опис генеративних архітектур, моделей, методів і даних, які були використані в роботі, також були визначені та описані методи аналізу розподілів даних в інформаційних представленнях генеративних моделей.

Четвертий розділ містить результати експериментальної частини роботи, які було розбито на три групи:

1. вимірювання та аналіз розподілів даних та структури генеративних представлень;

2. експериментальна перевірка методів неконтрольованого визначення структури щільності генеративних представлень (генеративного ландшафту представлень);

3. експериментальна перевірка методів навчання на генеративному ландшафті розроблених у роботі.

Отримані результати дають експериментальне підтвердження висновків теоретичної частини роботи про виникнення концептуальних структур представлень у процесі генеративного неконтрольованого навчання.

**Рекомендації щодо використання у дисертації результатів, одержаних автором.** Теоретичні та практичні результати дисертаційної роботи доцільно використовувати для розв'язання задач впевненого розпізнавання класів даних в сценаріях з мінімальними наборами навчальних даних, таких як дані трафіку Інтернет, що дозволить створити ефективні інтелектуальні системи для роботи в сфері інформаційної безпеки, громадської безпеки, медицині, а також при плануванні та прогнозуванні даних.

**Повнота викладення наукових положень дисертації в опублікованих працях.** Основні результати за темою дисертації автором викладено в 21 науковій праці, з них: 16 – у виданнях, що включені до міжнародної наукометричної бази Scopus, 1 – розділ в колективній монографії, 1 – у фаховому виданні України, 3 тези доповідей у міжнародних конференціях (Україна).

За своїм змістом та отриманими результатами дисертаційна робота відповідає паспорту спеціальності 05.13.06 – інформаційні технології, а саме: пункту 1 в частині розроблення «наукових і методологічних основ створення і застосування інформаційних технологій ... для автоматизованої переробки інформації і управління»; пункту 2 в частині розроблення «інформаційних технологій для аналізу та синтезу структурних, інформаційних ... моделей об'єктів і процесів, що автоматизуються», пункту 4 в частині побудова «інформаційних технологій для розроблення ... систем комп'ютерної підтримки рішень в автоматизованих системах і мережах». Вона є завершеною

кваліфікаційною працею з науковими положеннями, що надані автором для публічного захисту, характеризується внутрішньою єдністю та доводить особистий внесок автора в науку.

Зміст автореферату повністю відображає основні положення дисертаційної роботи.

**Зауваження до дисертації.** Незважаючи на достатній рівень виконаних наукових досліджень до дисертаційної роботи є такі зауваження:

1. В тексті та рисунках дисертаційної роботи автор використовує деякі нетипові (нестандартні) терміни.
2. У дисертаційній роботі опис наборів даних Інтернету (походження, зміст, процес отримання) може бути розширено за додатковими параметрами.
3. В розділі IV п. 4.1.1 розміри вибірки та кількість визначених латентних структур визначаються за допомогою методу зворотного виклику, бажано б було обґрунтувати використання цього методу та розглянути більш детально його застосування.
4. В дисертаційній роботі бажано більш детально описати специфічні умови використання запропонованої інформаційної технології, а також обмеження щодо сфер її застосування.
5. При подальшому дослідженні бажано навести конкретні приклади елементів набору даних Інтернет, як це було розглянуто при візуальному наборі даних.
6. У рефераті та дисертаційній роботі ілюстрації запропонованих методів навчання розпізнавання класів і концептів даних Інтернет можуть бути удосконалені для більшої інформативності; наприклад, детальний опис методів можна доповнити діаграмами чи схемами.



**Висновки.** Зазначені у відгуку зауваження не зменшують теоретичної та практичної цінності дисертаційної роботи ДОЛГИХ Сергія Миколайовича. Загалом, вона характеризується внутрішньою єдністю, виконана на належному науковому рівні та є завершеною працею. В ній отримано нові науково обґрунтовані результати, що в сукупності вирішують наукове завдання з розробки моделей та методів неконтрольованого глибокого навчання процесу розпізнавання даних пакетів трафіку Інтернет, на основі використання структури щільності генеративних представлень для розв'язання задачі впевненого розпізнавання категорій даних Інтернет у додатках, де значні масиви навчальних даних не доступні, або не можуть використовуватись внаслідок значної залежності успіху навчання традиційними методами від джерела даних. Дисертаційна робота Долгих С.М. відповідає вимогам «Порядку присудження наукових ступенів», а її автор заслуговує на присудження наукового ступеня кандидата технічних наук за спеціальністю 05.13.06 – інформаційні технології.

Офіційний опонент:

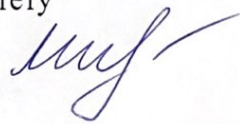
кандидат технічних наук, доцент,

доцент кафедри кібербезпеки та захисту інформації

факультету інформаційних технологій

Київського національного університету

імені Тараса Шевченка



Лариса МИРУТЕНКО

Підпис  
Секретар НДЧ  
Нарацька Н.В.  
18.05.2023р.

