УДК 004.8

В.І. КЛІМЕНКО, М.О. МОЛЧАНОВА, О.В. СОБКО,
В.І. АНДРОЩУК, О.В. МАЗУРЕЦЬ
Хмельницький національний університет

ПІДХІД ДО ПРОГРАМНОЇ ІНЖЕНЕРІЇ ТА ТЕСТУВАННЯ МЕТОДУ БАГАТОРІВНЕВОГО ВИЯВЛЕННЯ СУБ'ЄКТІВ КІБЕРБУЛІНГУ НА ОСНОВІ ТРАНСФОРМЕРІВ У ХМАРНОМУ СЕРЕДОВИЩІ

Стаття подає інженерно обґрунтований підхід до багаторівневого виявлення суб'єктів кібербулінгу з поєднанням первинної трансформерної детекції агресивних висловлювань та подальшої рольової інтерпретації дискурсу за схемою «ініціатор – мовленнєва дія – адресат» у середовищі Google Colab. Реалізація базується на попередньо натренованому класифікаторі cardiffnlp twitter roberta base offensive у конфігурації text classification бібліотеки transformers, що забезпечує відтворюваний інференс без додаткового донавчання і дає змогу зосередитися на програмній інженерії та тестуванні. Запропонована конструкція охоплює модулі інтерфейсу, керування обробкою, класифікації та синтаксико семантичного аналізу з фіксацією версій пакетів, випадкових зерен і параметрів токенізації, журнальним супроводом конфігурацій та збереженням артефактів на Google Drive. Методологічний внесок полягає у notebook oriented-організації життєвого циклу, що передбачає контроль формату даних, репліковані мікрозапуски, протокол порогування для бінарних рішень, маркування прикордонних випадків для ручного перегляду та вбудовані перевірки узгодженості рольової інтерпретації. Емпіричну дієздатність підтверджено на малій експертно верифікованій підвибірці, яка охоплює типові ситуації прямої образи, іронічного знецінення та нейтрального цитування потенційно токсичної лексики. Оцінювання включає фіксацію медіанного часу інференсу на прикладі міжквартильного розмаху як індикаторів операційної стабільності у змінних умовах колаб середовища, а також аналіз розподілу впевненості рішень з урахуванням робочого порога. Якісна валідація демонструє коректні спрацювання на явних інвективах, хибні позитиви у разі метамовного цитування та хибні негативи для саркастичних висловлювань, що обґрунтовує потребу рольової інтерпретації для зняття контекстної неоднозначності. Отримані результати свідчать, що поєднання попередньо натренованого класифікатора з рольовим рівнем і процедурно оформленим тестуванням забезпечує відтворюваність експериментів, інтерпретованість виходу і контрольовану детермінованість рішень у межах обчислювальних обмежень Colab, створюючи основу для масштабування на більші корпуси і багатомовні домени.

Ключові слова: *кібербулінг, рольова інтерпретація, трансформерні моделі, Google Colab.*

V.I. KLIMENKO, M.O. MOLCHANOVA, O.V. SOBKO,
V.I. ANDROSCHUK, O.V. MAZURETS
Khmelnytskyi National University

APPROACH TO SOFTWARE ENGINEERING AND TESTING OF THE METHOD OF MULTI-LEVEL DETECTION OF CYBERBULLYING SUBJECTS BASED ON TRANSFORMERS IN A CLOUD ENVIRONMENT

The article presents an engineering-based approach to multi-level detection of cyberbullying subjects with a combination of primary transformer detection of aggressive statements and subsequent role-based interpretation of discourse according to the initiator speech act addressee scheme in the Google Colab environment. The implementation is based on a pre-trained classifier cardiffnlp twitter roberta base offensive in the text classification configuration of the transformers library, which provides reproducible inference without additional training and allows you to focus on software engineering and testing. The proposed design includes interface modules, processing control, classification and syntactic semantic analysis with fixing package versions, random seeds and tokenization parameters, configuration logging and artifact storage on Google Drive. The methodological contribution consists of a notebook-oriented organization of the life cycle, which includes data format control, replicated microruns, a thresholding protocol for binary decisions, borderline case labeling for manual review, and built-in checks for consistency of role interpretation. Empirical feasibility is confirmed on a small expert-verified subsample, which covers typical situations of direct insult, ironic devaluation, and neutral citation of potentially toxic vocabulary. The evaluation includes fixing the median inference time per example and the interquartile range as indicators of operational stability in variable conditions of the collab environment, as well as analyzing the distribution of decision confidence taking into account the operating threshold. Qualitative validation demonstrates correct activations on explicit invectives, false positives in cases of metalinguistic citation, and false negatives for sarcastic statements, which justifies the need for role interpretation to remove contextual ambiguity. The results obtained indicate that the combination of a pre-trained classifier with a role level and procedurally designed testing provides reproducibility of experiments, interpretability of output, and controlled determinism of solutions within the computational constraints of Colab, creating a basis for scaling to larger corpora and multilingual domains.

Key words: *cyberbullying, role interpretation, transformative models, Google Colab.*

https://doi.org/10.32782/mathematical-modelling/2025-8-2-13 ISSN 2618-0332 (print), 2618-0340 (online)

Постановка проблеми

Ескалація агресивних взаємодій у цифрових середовищах дедалі частіше набуває непрямих форм з іронією, натяками, контекстними тригерами та короткими репліками-підсилювачами [1]. Більшість наявних рішень зупиняється на плоскому ярлику «токсично чи не токсично» і не встановлює адресності впливу, що унеможливлює змістовну модерацію [2]. Потрібен метод, який працює багаторівнево з первинним виявленням ознак кібербулінгу трансформерними моделями та подальшою інтерпретацією дискурсу через рольову конфігурацію «ініціатор – мовленнєва дія – адресат» із відокремленням підсилювачів і сторонніх учасників.

Поряд з алгоритмічним складником постає інженерне завдання забезпечити відтворюваність експериментів і перевірюваність якості в типових для дослідницької практики умовах Google Colab [3]. Сервіс Google Colab надає тимчасове середовище з мінливою конфігурацією, де сеанси перезапускаються, доступність і тип графічного прискорювача змінюються, а локальне сховище є непостійним. За таких умов життєвий цикл методу слід організувати як notebook driven engineering з фіксацією версій пакетів і випадкових зерен із детермінованими етапами завантаження та очищення даних, зі збереженням артефактів на Google Drive та з вбудованими мінітестами в ноутбуку для швидкої перевірки форматів і базових метрик [4].

Ключові виклики становлять мовна варіативність із жаргоном і кодміксом, орфографічні відхилення, дисбаланс класів і нестача явних міток адресності [5]. Необхідний конвеєр, що поєднує детекцію, рольову інтерпретацію та пояснюваність із локалізацією токсичних фрагментів, маркуванням ініціаторів і адресатів та контролем помилок високої вартості. Критерії якості мають охоплювати стабільність показників Precision Recall F1, поведінку порогів та латентність інференсу в межах сеансу Colab [6]. У підсумку проблема полягає у розробленні та верифікації підходу до програмної інженерії й тестування багаторівневого методу виявлення суб'єктів кібербулінгу на основі трансформерів у середовищі Google Colab, що забезпечує інтерпретовану рольову реконструкцію дискурсу, відтворювані процедури збереження і регресійної перевірки та операційну надійність вимірювань з урахуванням обмежень платформи.

Аналіз останніх досліджень і публікацій

У сучасному інформаційному просторі автоматизоване виявлення агресивних і токсичних висловлювань посідає ключове місце серед напрямів обробки природної мови, оскільки воно безпосередньо пов'язане з питаннями цифрової етики, модерації контенту та психологічної безпеки користувачів.

У роботах 2023–2024 рр. підтверджено ефективність донавчання сучасних трансформерів в мовах з обмеженими ресурсами. Для бенгальської мови моделі Bangla BERT і Multilingual BERT досягали F1 = 0.87 [7].

Новітні корпусні дослідження, присвячені автоматизованому аналізу агресивної комунікації, засвідчують еволюцію підходів від простої бінарної класифікації до багаторівневих і пояснювальних моделей. У межах конкурсу OffensEval, побудованого на таксономії OLID, продемонстровано, що архітектури типу трансформерів здатні не лише ефективно виявляти факти образливості, а й класифікувати їх за типом і визначати об'єкт спрямування. У першій ітерації (2019 р.) найвищі показники сягнули F1 = 0.829 для задачі A (детекція образливих висловлювань), F1 = 0.755 для задачі B (визначення, чи є образа таргетованою) та F1 = 0.660 для задачі C (ідентифікація мішені серед категорій IND/GRP/OTH), причому провідні результати показали моделі на базі BERT [8].

У сегменті коротких повідомлень на платформах мікроблогів спеціалізовані рішення на основі локальних твіттер-корпусів сягали F1 = 0.91 за умов цілеспрямованої доменної адаптації, що підкреслює її критичне значення [9].

Пояснювальний вимір проблеми було поглиблено в рамках SemEval 2021 Task 5 – Toxic Spans, де завданням стало точне виділення токсичних елементів тексту на рівні символів. Найуспішніша модель досягла показника character F1 = 70.83%, що засвідчило реалістичність підходів до раціоналізації рішень трансформерними моделями. Водночас спостерігалася суттєва розбіжність у продуктивності: системи, що використовували BiLSTM-CRF або ToxicBERT,

https://doi.org/10.32782/mathematical-modelling/2025-8-2-13 ISSN 2618-0332 (print), 2618-0340 (online)

показували F1 близько 62.23%, що вказує на складність задачі спан-детекції та високу залежність результатів від архітектури й параметризації [10].

Загалом динаміка розвитку завдань у сфері виявлення кібербулінгу демонструє поступову інтеграцію контекстно-чутливих моделей із пояснювальними механізмами. Сучасні трансформери довели свою здатність не лише класифікувати кібербулінг, а й надавати інтерпретовані результати, що сприяє підвищенню довіри до систем автоматизованої модерації та розширює можливості їх застосування у багатомовному середовищі.

Мета дослідження

Мета полягає в обґрунтуванні та реалізації цілісного підходу до програмної інженерії і перевірки якості багаторівневого методу виявлення суб'єктів кібербулінгу на основі трансформерів у середовищі Google Colab із зосередженням на рольовій інтерпретації дискурсу за схемою «ініціатор – мовленнєва дія – адресат». Дослідження спрямоване на створення відтворюваного ноутбук-орієнтованого конвеєра з фіксацією версій і контрольованою детермінованістю, на формування процедурного каркаса тестування, що охоплює перевірки коректності даних, стабільність метрик точності та повноти, поведінку порогів і латентність інференсу в межах обчислювальних обмежень Colab, а також на розроблення інженерних артефактів, придатних для повторної збірки і незалежної валідації. Реалізація методу здійснюється на базі попередньо натренованого класифікатора cardiffnlp/twitter-roberta-base-offensive [11] у конфігурації text classification із використанням бібліотеки transformers у середовищі Google Colab, що забезпечує відтворюваність та уніфікованість інференсу без додаткового донавчання моделі.

Виклад основного матеріалу дослідження

У роботі розглянуто метод багаторівневого виявлення суб'єктів кібербулінгу, який поєднує первинну трансформерну детекцію агресивних висловлювань із подальшою рольовою інтерпретацією дискурсу за схемою «ініціатор – мовленнєва дія – адресат». Схему методу наведено на рис. 1. На етапі первинної детекції використано попередньо натреновану модель cardiffnlp/twitter-roberta-base-offensive як базовий компонент агресивного мовлення. Такий вибір дає змогу зосередитися на інженерних аспектах інтеграції і тестування багаторівневого конвеєра без витрат на навчання з нуля. Мотиваційну базу становлять сучасні корпусні ініціативи, що переходять від бінарної класифікації до задач з ідентифікацією мішені та виділенням токсичних фрагментів, що демонструє практичну реалізованість спан-орієнтованого раціоналізування рішень на базі трансформерів і задає вимоги до інтерпретованості моделі у прикладних сценаріях модерації.

Запропонований підхід позиціонується як інженерно завершене рішення, у якому трансформерна детекція та рольова інтерпретація інтегровані зі специфікаціями даних, шаблоном експерименту та контрольованим тестовим контуром у Google Colab. Така постановка забезпечує не лише якісні показники класифікації, а й перевірюваність відтворюваних запусків, що безпосередньо відповідає акценту статті на програмній інженерії та тестуванні методу.

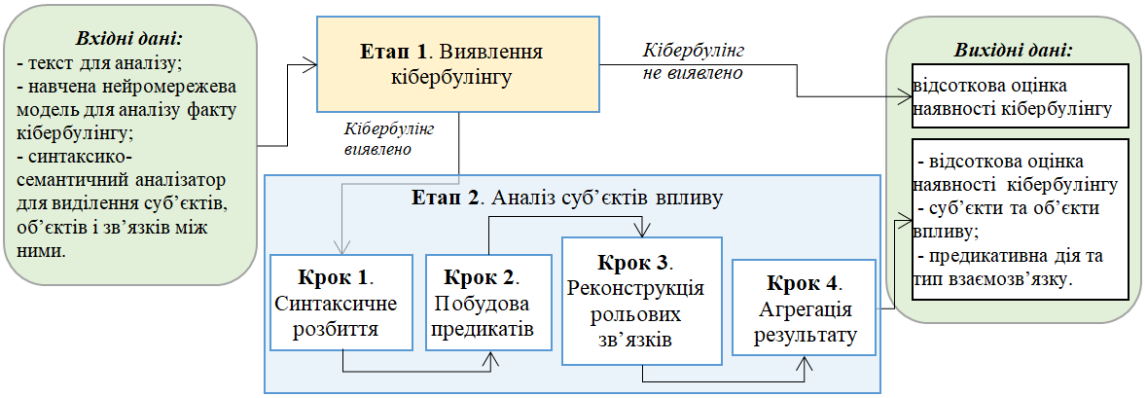

Рис. 1. Схема методу багаторівневого виявлення суб'єктів кібербулінгу

https://doi.org/10.32782/mathematical-modelling/2025-8-2-13 ISSN 2618-0332 (print), 2618-0340 (online)

Архітектурна організація реалізації ґрунтується на хмарному середовищі Google Colab як на дослідницькому контейнері для розроблення і запуску. Усередині середовища функціонує Python Runtime з модулями інтерфейсу, керування обробкою, трансформерної класифікації та синтаксико семантичного аналізу. Завантаження моделей здійснюється з HuggingFace Hub, а артефакти зберігаються на Google Drive, що забезпечує відтворюваність і перенесення між сесіями [12]. Візуальне представлення взаємодії компонентів наведено на рис. 2, де зафіксовано ключові залежності між інтерфейсом Gradio, контролером виконання, трансформерними моделями та інструментами лінгвістичного аналізу. Архітектурна схема охоплює шари підготовки даних, інференсу та інтерпретації результатів із журналюванням експериментів у вигляді JSON-логу та табличних підсумків, що зберігаються на Google Drive разом із контрольними знімками версій пакетів і випадкових зерен. Така організація робить можливими регресійні запуски в ідентичних умовах наступних сесій Colab та мінімізує вплив апаратних флуктуацій. Базовий класифікаційний модуль реалізовано через стандартний пайплайн transformers із фіксацією версій пакетів, параметрів токенізації та хешу ваг моделі. Усі артефакти інференсу, журнали конфігурації та контрольні підвибірки зберігаються на Google Drive, що уможливлює повторення запусків у наступних сесіях Colab в ідентичних умовах.

Логіка даних і результатів організована довкола узгоджених артефактів прогнозування і рольової інтерпретації. Імовірнісний вихід класифікатора та бінарне рішення після порогування інкапсульовано в структурі прогнозу, тоді як результат запиту агрегує пов'язані об'єкти рольової інтерпретації та тріади учасників, що походять із синтаксичних структур, після чого повертається до інтерфейсу користувача. Така композиція об'єктів відповідає конвеєру від текстового введення через аналіз залежностей до формування інтерпретованого результату й узгоджується з архітектурною схемою, що наведена на рис. 2, а також із конфігурацією попередньо натренованого класифікатора, який забезпечує стабільне перетворення тексту в імовірнісні оцінки.

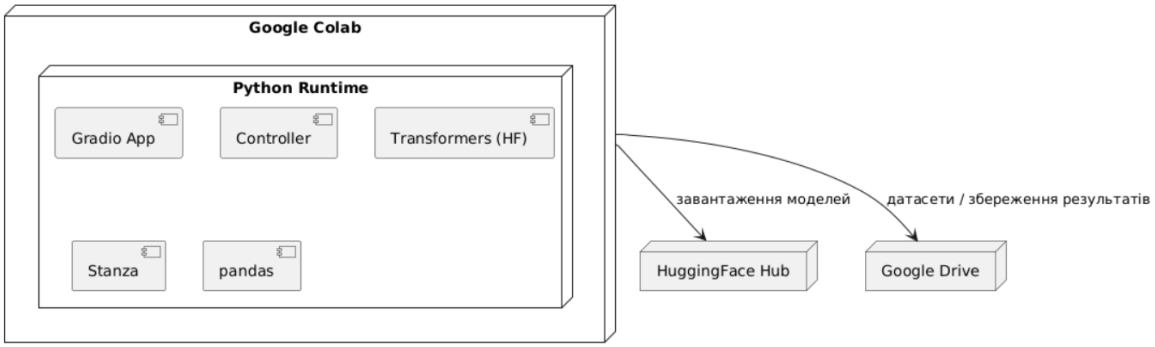

Рис. 2. Взаємодія компонентів і хмарного середовища

Рольовий підхід спирається на дискурсивне моделювання учасників взаємодії з фіксацією ініціатора і адресата, а також урахуванням спостерігачів підсилювачів і захисників. У мережевих середовищах функціональні ролі є динамічними і залежать від локального контексту та ходу обговорення, що безпосередньо мотивує багаторівневу постановку задачі з рівнями повідомлення, діалогу та взаємодії. Запропонована інтерпретація уточнює предметну сферу методу і визначає вимоги до модулів.

Рольова інтерпретація безпосередньо реалізує заявлений у назві фокус багаторівневого виявлення, оскільки поєднує повідомленнєвий рівень виявлення з дискурсивним рівнем адресності. У результаті отримується структурований вихід, який придатний до аналітичного використання у модераційних сценаріях і підлягає перевірці стандартними інженерними

https://doi.org/10.32782/mathematical-modelling/2025-8-2-13 ISSN 2618-0332 (print), 2618-0340 (online)

процедурами тестування. З огляду на використання попередньо натренованого класифікатора, інженерна новизна полягає у конструкції рольового рівня, протоколів відтворюваності та перевірок стабільності інференсу в Colab.

Тестовий контур сформовано у межах ноутбука Colab і спрямовано на процедурну перевірку попередньо натренованого класифікатора та узгодженості рольового рівня. Перевірка коректності охоплює валідацію форматів вхідних даних, відтворюваність вихідних оцінок за повторних проходів тих самих прикладів, чутливість до емодзі, орфографічних відхилень і коди міксу, а також санітарні приклади з нейтральними, позитивними та завуальовано агресивними репліками. Для кількісної фіксації придатні базові показники точності й повноти на малій експертно верифікованій підвибірці без обов'язкового включення повної матриці помилок у текст статті. Додатково оцінюється латентність інференсу як медіанний час, наприклад із міжквартильним розмахом, що дає змогу врахувати варіації ресурсів Colab і підкріплює твердження про операційну надійність.

Робочий поріг для бінарного рішення фіксується у протоколі експерименту, а приклади з низькою впевненістю моделі позначаються як прикордонні для подальшого ручного перегляду. Такий протокол знижує ризик хибних рішень у прикладних модераційних сценаріях і забезпечує відтворюваність тонкого налаштування без донавчання моделі.

Алгоритмічна частина реалізує відображення від тексту до рішення про наявність проявів кібербулінгу та до пов'язаних із цим рольових тріад. Для підвищення пояснюваності застосовується локалізація індикативних фрагментів, що корелює з тенденціями у спан-орієнтованих завданнях і допускає подальше поєднання із залежнісним синтаксичним аналізом для відновлення предикативних зв'язків між учасниками і діями, що їх пов'язують. Така конструкція надає можливість аналізувати як явні прямі агресивні висловлювання, так і непрямі мовленнєві дії у стислих репліках і контекстуально забарвлених відповідях.

Приклад роботи програмного продукту для багаторівневого виявлення суб'єктів кібербулінгу наведено на рис. 3.

Представлена конструкція охоплює повний цикл – від архітектури до тестування у реаліях Google Colab і підтверджує заявлену в назві орієнтацію на програмну інженерію та перевірку якості методу. Поєднання трансформерної детекції, рольової інтерпретації та інструментованого тестового контуру дає відтворюваний результат із контрольованою детермінованістю та інтерпретованим виходом, що уможливлює незалежну валідацію й перенесення підходу на суміжні домени комунікації.

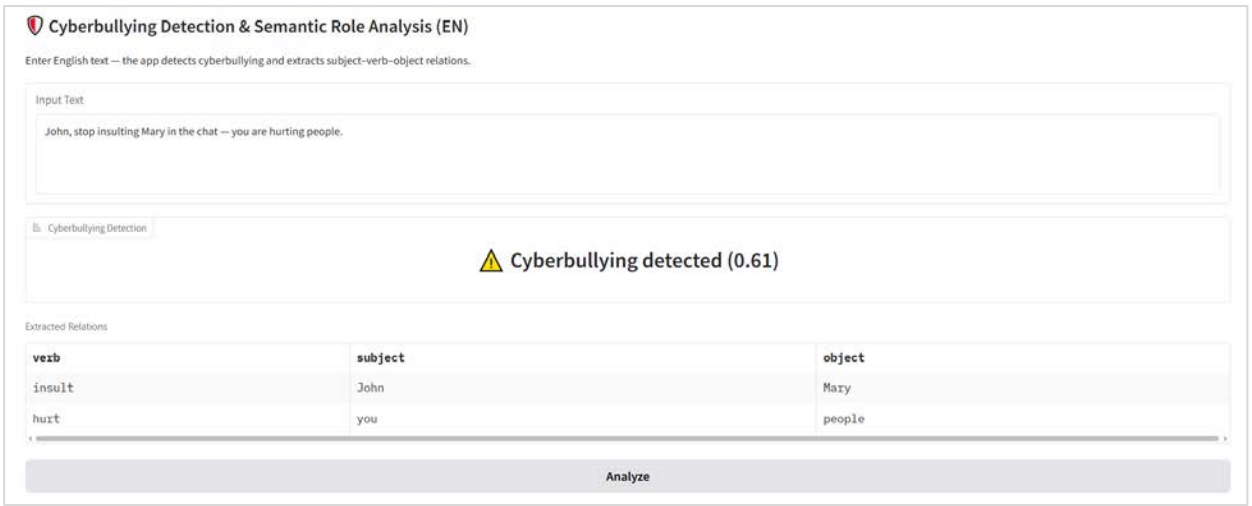

Рис. 3. Приклад роботи створеного програмного застосунку

https://doi.org/10.32782/mathematical-modelling/2025-8-2-13 ISSN 2618-0332 (print), 2618-0340 (online)

Дієздатність конвеєра підтверджено відтворюваними запусками у Google Colab із фіксацією версій пакетів, випадкових зерен і конфігурації пайплайна. Перевірку проведено на невеликій експертно верифікованій підвибірці, що віддзеркалює типові для цифрової комунікації ситуації прямої образи, іронічного знецінення та нейтральних цитат із потенційно токсичним лексиконом. Попередньо натренований класифікатор cardiffnlp/twitter-roberta-base-offensive стабільно формує ймовірнісні оцінки для кожного прикладу, після чого бінарне рішення приймається за зафіксованим робочим порогом, зазначеним у протоколі експерименту. Для фіксації поведінки інференсу подається медіанний час обробки і міжквартильний розмах як показник стабільності в межах змінної доступності обчислювальних ресурсів Colab. Ілюстративний розподіл упевненості рішень для підвибірки наведено на рис. 4.

Для підтвердження практичної корисності наведено якісну валідацію на підібраних прикладах із контрастним контекстом. Приклади прямої адресної образи зазвичай супроводжуються високою впевненістю класифікатора і послідовним бінарним рішенням, що відтворюється за повторних проходів. Приклади нейтральних повідомлень із цитуванням токсичної лексики у метамовній функції демонструють очікувану нейтральну класифікацію за умови консервативного робочого порогу, що підкреслює значущість протоколу порогування для операційного застосування.

Найбільш показовими виявляються прикордонні іронічні висловлювання та короткі репліки з контекстним знеціненням, де значення впевненості наближаються до порогу. У таких випадках рішення маркується як прикордонне і підлягає ручному перегляду з урахуванням рольової інтерпретації за схемою «ініціатор – мовленнєва дія – адресат». Саме рольовий рівень знімає неоднозначність, оскільки дає змогу зіставити висловлювання з локальним дискурсом, виокремити спрямованість і зафіксувати, чи є адресат очевидним або імпліцитним.

Для прозорості викладу наведено зведену табл. 1 ілюстративних прикладів, де для кожного рядка вказано текст фрагмента, оцінку впевненості моделі, бінарне рішення, експертне судження та короткий коментар щодо природи помилки або успішної класифікації. Така форма відповідає вимогам відтворюваного тестування: таблиця може бути доповнена новими прикладами без зміни протоколу, а прикордонні випадки залишаються доступними для подальшої ручної перевірки та узгодження порога.

Рис. 4. Розподіл упевненості класифікатора для контрольної підвибірки

Таблиця 1

Приклади коректних і хибних класифікацій із коментарями експерта

№	Текст фрагмента	Упевненість моделі	Рішення моделі	Експертне судження	Коментар (ініціатор – мовленнєва дія – адресат)
1	You are such an idiot!	0.98	Offensive	Offensive	Ініціатор здійснює пряму образу адресата; висока впевненість, узгоджене рішення
2	That guy is a total loser	0.91	Offensive	Offensive	Пряма адресна зневага; рольова конфігурація є очевидною
3	Shut up already	0.88	Offensive	Offensive	Наказова форма знижує ввічливість, явна агресія
4	He said “you are an idiot”, which is unacceptable	0.36	Not offensive	Not offensive	Метамовне цитування без власної агресії; контекст нейтральний
5	Quoting: “moron” — this word hurts	0.29	Not offensive	Not offensive	Опис наслідків лексеми, а не її вживання як інвективи

https://doi.org/10.32782/mathematical-modelling/2025-8-2-13 ISSN 2618-0332 (print), 2618-0340 (online)

Продовження таблиці 1

№	Текст фрагмента	Упевненість моделі	Рішення моделі	Експертне судження	Коментар (ініціатор – мовленнєва дія – адресат)
6	I think the post is stupid, maybe read before posting	0.54	Offensive	Borderline	Іронічне знецінення; близько до порогу, потребує рольової перевірки контексту
7	Great, another genius move…	0.47	Not offensive	Borderline	Сарказм без явної адресності; прикордонний випадок
8	You “smart” as always	0.52	Offensive	Borderline	Саркастичний епітет може мати приховану адресність; до ручного перегляду
9	This is nonsense	0.4	Not offensive	Not offensive	Критика ідеї без персональної адресності
10	You’re pathetic	0.93	Offensive	Offensive	Персональна образа з високою впевненістю

Дієздатність і межі застосовності методу продемонстровано на прикладах табл. 1, а також на розподілі впевненості для контрольної підвибірки, поданому на рис. 4.

Висновки

Дослідження продемонструвало практичну придатність інженерно організованого підходу до багаторівневого виявлення суб’єктів кібербулінгу у середовищі Google Colab із поєднанням трансформерної детекції та рольової інтерпретації дискурсу за схемою «ініціатор – мовленнєва дія – адресат». Використання попередньо натренованого класифікатора cardiffnlp twitter roberta base offensive у конфігурації text classification дало змогу сконцентруватися на відтворюваному життєвому циклі методу та на процедурному тестуванні без додаткового донавчання. Сформовано ноутбук-орієнтований конвеєр із фіксацією версій і випадкових зерен, контрольованими мікрозапусками, протоколом порогування бінарних рішень та маркуванням прикордонних випадків для ручного перегляду, що забезпечує стабільність інференсу та інтерпретованість виходу в умовах змінних ресурсів Colab.

Емпірична перевірка на малій експертно верифікованій підвибірці підтвердила, що метод відтворювано розрізняє явні інвективи, нейтральні повідомлення з метамовним цитуванням та іронічні висловлювання, а рольовий рівень знімає частину контекстної неоднозначності у прикордонній зоні рішень. Фіксація медіанного часу інференсу та його варіативності засвідчила операційну надійність конвеєра у межах обмежень платформи, а протокол відтворюваності дає можливість незалежної валідації результатів у наступних сесіях.

Обмеження дослідження пов’язані з невеликим обсягом контрольної вибірки, відсутністю донавчання під конкретний домен і залежністю від англомовної моделі, що зумовлює чутливість до коди міксу та саркастичних формулювань. Подальший розвиток слід спрямувати на розширення корпусів і багатомовних доменів, калібрування впевненості та уточнення порогів, інтеграцію напівавтоматичного узгодження рольових рішень із залученням експертів, а також на оцінювання переносимості підходу у сценаріях із різною жанровою структурою та різним ступенем адресності. Отримані результати підтверджують доцільність інженерної парадигми відтворюваного тестування у Colab як основи для масштабування і подальшого підвищення якості багаторівневого виявлення суб’єктів кібербулінгу.

Список використаної літератури

1. Wang S., Shibghatullah A.S., Iqbal T.J. et al.. A review of multimodal-based emotion recognition techniques for cyberbullying detection in online social media platforms. *Neural Computing and Applications.* 2024. № 36. P. 21923–21956. https://doi.org/10.1007/s00521-024-10371-3

2. Islam M.M., Uddin M.A., Islam L., Akter A., Sharmin S., Acharjee U.K. Cyberbullying detection on social networks using machine learning approaches. 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE), December 2020. International Journal of Computer Applications. January 2025. Vol. 186, No. 61. P. 1–6. IEEE. https://doi.org/10.5120/ijca2025924395
3. Молчанова М.О., Дідур В.О., Мазурець О.В., Тищенко О.О., Залуцька О.О. Інформаційна технологія використання хмарних обчислень для класифікації залишків зруйнованих будівель засобами нейронних мереж за візуальними даними з безпілотних літальних апаратів. *Наука і техніка сьогодні.* 2025. № 4 (45). С. 1259–1272. https://doi.org/10.52058/2786-6025-2025-4(45)-1259-1272
4. Молчанова М.О., Мазурець О.В., Собко О.В., Кліменко В.І., Андрощук В.І. Метод нейромережевого виявлення кібербулінгу з використанням хмарних сервісів та об'єктно-орієнтованої моделі. Вісник Хмельницького національного університету. Серія «Технічні науки». 2024. № 2(333). С. 200–206. https://doi.org/10.31891/2307-5732-2024-333-2
5. Chen M., et al. 1st Place Solution to Odyssey Emotion Recognition Challenge Task1: Tackling Class Imbalance Problem. The Speaker and Language Recognition Workshop (Odyssey 2024), 18–21 June 2024, Quebec City, Canada. ISCA, 2024. P. 1–6. https://doi.org/10.21437/odyssey.2024-37
6. Yin Y., Feng Y., Weng S., Gao X., Liu J., Zhao Z. Lightweight Probabilistic Coverage Metrics for Efficient Testing of Deep Neural Networks. Proceedings of the 16th International Conference on Internetware (Internetware '25), June 20–22, 2025. Association for Computing Machinery, New York, NY, USA. P. 474–486. https://doi.org/10.1145/3755881.3755915
7. Sihab-Us-Sakib S., Rahman M.R., Forhad M.S.A., Aziz M.A. Cyberbullying detection of resource-constrained language from social media using transformer-based approach. Natural Language Processing Journal. 2024. Vol. 9. P. 100104. https://doi.org/10.1016/j.nlp.2024.100104
8. Zampieri M., Rosenthal S., Nakov P., Dmonte A., Ranasinghe T. OffensEval 2023: Offensive language identification in the age of Large Language Models. Natural Language Engineering. 2023. Vol. 29, No. 6. P. 1416–1435. https://doi.org/10.1017/S1351324923000517
9. Aliyeva Ç.O., Yağanoğlu M. Deep learning approach to detect cyberbullying on Twitter. Multimedia Tools and Applications. 2024. Vol. 84. P. 20497–20520. https://doi.org/10.1007/s11042-024-19869-3
10. Pavlopoulos J., Sorensen J., Laugier L., Androutsopoulos I. SemEval-2021 Task 5: Toxic Spans Detection. Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021), Online, August 2021. Association for Computational Linguistics. P. 59–69. https://doi.org/10.18653/v1/2021.semeval-1.6
11. Twitter-roBERTa-base for Offensive Language Identification: *Huggingface* URL: https://huggingface.co/cardiffnlp/twitter-roberta-base-offensive (дата звернення: 10.11.2025).
12. Krak I., Molchanova M., Didur V., Sobko O., Mazurets O., Barmak O. Method of semantic features estimation for political propaganda techniques detection using transformer neural networks. CEUR Workshop Proceedings, Vol. 3917. Kryvyi Rih, Ukraine, December 27, 2024. P. 286–297. URL: https://ceur-ws.org/Vol-3917/paper56.pdf (дата звернення: 10.11.2025).

References

1. Wang, S., Shibghatullah, A.S., Iqbal, T.J., et al. (2024). A review of multimodal-based emotion recognition techniques for cyberbullying detection in online social media platforms. *Neural Computing and Applications,* 36, 21923–21956. https://doi.org/10.1007/s00521-024-10371-3 [in English].
2. Islam, M.M., Uddin, M.A., Islam, L., Akter, A., Sharmin, S., & Acharjee, U.K. (2020, December). Cyberbullying detection on social networks using machine learning approaches. Proceedings of

the 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE). International Journal of Computer Applications, 186(61), 1–6, January 2025. IEEE. https://doi.org/10.5120/ijca2025924395 [in English].

3. Molchanova, M.O., Didur, V.O., Mazurets, O.V., Tyshchenko, O.O., & Zalutska, O.O. (2025). Informatsiina tekhnolohiia vykorystannia khmarnykh obchyslen dlia klasyfikatsii zalyshkiv zruinovanykh budivel zasobamy neironnykh merezh za vizualnymy danymy z bezpilotnykh litalnykh aparativ [Information technology of cloud computing use for classification of destroyed buildings' remnants by neural networks based on UAV visual data]. Naukovyi zhurnal «Nauka i tekhnika sohodni», 4(45), 1259–1272. https://doi.org/10.52058/2786-6025-2025-4(45)-1259-1272 [in Ukrainian].
4. Molchanova, M.O., Mazurets, O.V., Sobko, O.V., Klimenko, V.I., & Androshchuk, V.I. (2024). Metod neiromerezhevoho vyiavlennia kiberbulinhu z vykorystanniam khmarnykh servisiv ta obiektno-oriientovanoi modeli [Neural network method for cyberbullying detection using cloud services and object-oriented model]. Naukovyi zhurnal «Visnyk Khmelnytskoho natsionalnoho universytetu». Seriia: Tekhnichni nauky, 2(333), 200–206. https://doi.org/10.31891/2307-5732-2024-333-2 [in Ukrainian].
5. Chen, M., et al. (2024). 1st Place Solution to Odyssey Emotion Recognition Challenge Task1: Tackling Class Imbalance Problem. In The Speaker and Language Recognition Workshop (Odyssey 2024), Quebec City, Canada (pp. 1–6). ISCA. https://doi.org/10.21437/odyssey.2024-37 [in English].
6. Yin, Y., Feng, Y., Weng, S., Gao, X., Liu, J., & Zhao, Z. (2025). Lightweight Probabilistic Coverage Metrics for Efficient Testing of Deep Neural Networks. In Proceedings of the 16th International Conference on Internetware (Internetware '25) (pp. 474–486). Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3755881.3755915 [in English].
7. Sihab-Us-Sakib, S., Rahman, M.R., Forhad, M.S.A., & Aziz, M.A. (2024). Cyberbullying detection of resource-constrained language from social media using transformer-based approach. Natural Language Processing Journal, 9, 100104. https://doi.org/10.1016/j.nlp.2024.100104 [in English].
8. Zampieri, M., Rosenthal, S., Nakov, P., Dmonte, A., & Ranasinghe, T. (2023). OffensEval 2023: Offensive language identification in the age of Large Language Models. Natural Language Engineering, 29(6), 1416–1435. https://doi.org/10.1017/S1351324923000517 [in English].
9. Aliyeva, Ç.O., & Yağanoğlu, M. (2024). Deep learning approach to detect cyberbullying on Twitter. Multimedia Tools and Applications, 84, 20497–20520. https://doi.org/10.1007/s11042-024-19869-3 [in English].
10. Pavlopoulos, J., Sorensen, J., Laugier, L., & Androutsopoulos, I. (2021, August). SemEval-2021 Task 5: Toxic Spans Detection. Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021), Online (pp. 59–69). Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.semeval-1.6 [in English].
11. Twitter-roBERTa-base for offensive language identification. (2025) *Hugging Face*. Retrieved from https://huggingface.co/cardiffnlp/twitter-roberta-base-offensive.
12. Krak, I., Molchanova, M., Didur, V., Sobko, O., Mazurets, O., & Barmak, O. (2025, December 27). Method of semantic features estimation for political propaganda techniques detection using transformer neural networks. In CEUR Workshop Proceedings (Vol. 3917, pp. 286–297). Virtual Event, Kryvyi Rih, Ukraine. Retrieved from https://ceur-ws.org/Vol-3917/paper56.pdf [in English].

Кліменко Валерія Ігорівна – асистент кафедри комп'ютерних наук Хмельницького національного університету. E-mail: ler.klimenko.8@gmail.com, ORCID: 0000-0001-5869-4269.

Молчанова Марина Олексіївна – доктор філософії з комп'ютерних наук, старший викладач кафедри комп'ютерних наук Хмельницького національного університету. E-mail: m.o.molchanova@gmail.com, ORCID: 0000-0001-9810-936X.

Собко Олена Віталіївна – доктор філософії з комп'ютерних наук, старший викладач кафедри комп'ютерних наук Хмельницького національного університету. E-mail: olenasobko.ua@gmail.com, ORCID: 0000-0001-5371-5788.

Андрощук Владислав Іванович – студент кафедри комп'ютерних наук Хмельницького національного університету. E-mail: vladandroschuk0@gmail.com, ORCID: 0009-0006-1910-7221.

Мазурець Олександр Вікторович – к.т.н., доцент, доцент кафедри комп'ютерних наук Хмельницького національного університету. E-mail: exe.chong@gmail.com, ORCID: 0000-0002-8900-0650.

Klimenko Valeria Ihorivna – Assistant Professor at the Department of Computer Science of the Khmelnytskyi National University. E-mail: ler.klimenko.8@gmail.com, ORCID: 0000-0001-5869-4269.

Molchanova Maryna Oleksiivna – Ph.D. in Computer Science, Senior Lecturer at the Department of Computer Science of the Khmelnytskyi National University. E-mail: m.o.molchanova@gmail.com, ORCID: 0000-0001-9810-936X.

Sobko Olena Vitaliivna – Ph.D. in Computer Science, Senior Lecturer at the Department of Computer Science of the Khmelnytskyi National University. E-mail: olenasobko.ua@gmail.com, ORCID: 0000-0001-5371-5788.

Androshchuk Vladyslav Ivanovych – Student at the Department of Computer Science of the Khmelnytskyi National University. E-mail: vladandroschuk0@gmail.com, ORCID: 0009-0006-1910-7221

Mazurets Oleksandr Viktorovych – Candidate of Technical Sciences, Associate Professor, Associate Professor at the Department of Computer Science of the Khmelnytskyi National University. E-mail: exe.chong@gmail.com, ORCID: 0000-0002-8900-0650.

Отримано: 31.10.2025
Рекомендовано: 28.11.2025
Опубліковано: 30.12.2025