УДК 004.8
DOI https://doi.org/10.32782/2521-6643-2026-1-71.10

**Молчанова М. О.,** доктор філософії,
старший викладач кафедри комп'ютерних наук
Хмельницького національного університету
ORCID: 0000-0001-9810-936X

**Андрощук В. І.,** здобувач вищої освіти кафедри комп'ютерних наук
Хмельницького національного університету
ORCID: 0009-0006-1910-7221

**Шурипа М. О.,** здобувач вищої освіти кафедри комп'ютерних наук
Хмельницького національного університету
ORCID: 0009-0003-7025-4647

**Мазурець О. В.,** кандидат технічних наук, доцент,
доцент кафедри комп'ютерних наук
Хмельницького національного університету
ORCID: 0000-0002-8900-0650

# ОБ'ЄКТНО-ОРІЄНТОВАНИЙ ПІДХІД ДО НЕЙРОМЕРЕЖЕВОГО ВИЯВЛЕННЯ СУБ'ЄКТІВ КІБЕРБУЛІНГУ ЗА ПОВІДОМЛЕННЯМИ У КЕРОВАНОМУ ХМАРНОМУ СЕРЕДОВИЩІ

*Метою роботи є формування та обґрунтування об'єктно-орієнтованого підходу до нейромережевого виявлення суб'єктів кібербулінгу за повідомленнями із поєднанням первинної детекції та подальшої синтаксико-семантичної інтерпретації у керованому хмарному середовищі. Запропоновано узгоджену архітектуру, у якій нейромережевий модуль здійснює фільтрацію повідомлень на рівні «кібербулінг / не кібербулінг», після чого результати проходять залежнісний аналіз із реконструкцією рольових зв'язків типу «суб'єкт – дія – об'єкт». Об'єктну модель, що включає класи повідомлень, речень, токенів, залежностей, предикатів, рольових трійок і підсумкових структур, визначено як основу для забезпечення прозорої трасованості рішень, тоді як кероване хмарне середовище забезпечує відтворюваність запусків та масштабованість експериментів.*

*Експериментально підтверджено ефективність первинної детекції: модуль на основі BERT продемонстрував метрику F1 = 0.98 у двокласовій постановці («кібербулінг» / «не кібербулінг»), що свідчить про достатній рівень відсіву нерелевантних повідомлень перед рольовим аналізом. На експертно верифікованому піднаборі встановлено узгоджені показники якості рольової ідентифікації: для суб'єкта отримано значення 0.88, 0.86, 0.87 за Precision, Recall і F1 відповідно; для об'єкта – 0.85, 0.83, 0.84; для дієслівного центру – 0.91, 0.89, 0.90. Точне відновлення рольової трійки забезпечило значення F1 = 0.76. Міжекспертна узгодженість становила коефіцієнт Коена 0.82 при 87 % повної згоди, що засвідчує надійність еталонних міток і коректність застосованої процедури оцінювання. У спірних випадках використано третю допоміжну оцінку мовною моделлю із фіксованою інструкцією; фінальні мітки визначено за правилом більшості.*

*Отримано результати, які демонструють, що запропонований підхід не лише дає змогу визначити факт агресивної комунікації, а й забезпечує структуроване подання інформації про її адресність, залишаючись відтворюваним та аудитованим у практичних умовах. Сформовано висновки, що створюють підґрунтя для подальшої інтеграції підходу у модераційні системи й можливого розширення на корпуси з детальнішою рольовою розміткою та багатомовною підтримкою.*

Ключові слова: *кібербулінг, трансформерні моделі, синтаксико-семантичний аналіз, об'єктно-орієнтоване проектування, кероване хмарне середовище.*

***Molchanova M. O., Androshchuk V. I., Shurypa M. O., Mazurets O. V. Object-oriented approach to neural network-based detection of cyberbullying subjects from messages in a managed cloud environment***

*The aim of the work is to develop and substantiate an object-oriented approach to neural network detection of cyberbullying subjects from messages with a combination of primary detection and subsequent syntactic-semantic interpretation in a managed cloud environment. A coherent architecture is proposed in which the neural network module filters messages at the "cyberbullying / non-cyberbullying" level, after which the results undergo dependency analysis with the reconstruction of role relationships of the "subject – action – object" type. An object model, which includes classes of messages, sentences, tokens,*

*dependencies, predicates, role triples and summary structures, is defined as the basis for ensuring transparent traceability of decisions, while a managed cloud environment ensures the reproducibility of launches and scalability of experiments.*

*The effectiveness of the initial detection was experimentally confirmed: the BERT-based module demonstrated a metric of F1 = 0.98 in a two-class setting ("cyberbullying" / "not cyberbullying"), which indicates a sufficient level of screening out irrelevant messages before role analysis. Consistent indicators of the quality of role identification were established on the expert-verified subset: for the subject, values of 0.88, 0.86, 0.87 were obtained for Precision, Recall and F1, respectively; for the object – 0.85, 0.83, 0.84; for the verb center – 0.91, 0.89, 0.90. The exact restoration of the role triple provided a value of F1 = 0.76. The inter-expert agreement was Cohen's coefficient of 0.82 with 87 % complete agreement, which indicates the reliability of the reference labels and the correctness of the applied evaluation procedure. In controversial cases, a third auxiliary assessment by a language model with a fixed instruction was used; the final labels were determined by majority rule.*

*The results obtained demonstrate that the proposed approach not only allows to determine the fact of aggressive communication, but also provides a structured presentation of information about its addressability, remaining reproducible and auditable in practical conditions. Conclusions are drawn that create a basis for further integration of the approach into moderation systems and possible expansion to corpora with more detailed role markup and multilingual support.*

Key words: *cyberbullying, transformer models, dependency parsing, object-oriented design, managed cloud.*

**Постановка проблеми.** У сучасних цифрових комунікаціях кібербулінг набуває контекстно залежних і часто непрямих форм [1]. Переважна частина автоматизованих рішень обмежується бінарною класифікацією повідомлень на «агресивні/неагресивні», що не забезпечує встановлення адресності впливу – зокрема, ідентифікації ініціатора, цільової особи та характеру мовленнєвого акту [2]. Відсутність рольової інтерпретації ускладнює побудову адресних інтервенцій, аудит рішень і аналітичний супровід модерації [3].

Актуальною є науково-практична задача розроблення багаторівневого підходу, який поєднує визначення наявності кібербулінгу [4] із подальшою реконструкцією суб'єктно-об'єктних зв'язків у висловлюваннях [5]. З інженерного погляду така задача потребує чіткої об'єктно-орієнтованої моделі предметної області, у якій повідомлення, речення, предикати та рольові трійки репрезентуються як окремі сутності із визначеними відповідальностями та інваріантами. Об'єктно-орієнтований підхід покликаний забезпечити модульність, розширюваність і тестованість багаторівневої обробки [6], від нейромережевої детекції ознак кібербулінгу до синтаксико-семантичного аналізу й інтерпретації ролей учасників.

Додаткові вимоги висуває обчислювальний контекст: результати мають бути відтворюваними та масштабованими у керованому хмарному середовищі, придатному для експериментів і прискореного інференсу [7]. Отже, постає потреба в методі та програмній реалізації, які поєднують нейромережеву детекцію з лінгвістично вмотивованою рольовою інтерпретацією, спираються на об'єктно-орієнтоване моделювання домену й забезпечують умови для подальшого оцінювання якості, продуктивності та практичної придатності.

**Аналіз останніх досліджень і публікацій.** Сучасні корпусні ініціативи з автоматизованого аналізу агресивної комунікації демонструють перехід від бінарної детекції до ієрархічних та пояснювальних постановок. У межах OffensEval на базі таксономії OLID показано, що трансформерні моделі ефективні не лише для виявлення образливості, а й для подальшої категоризації та ідентифікації мішені. У першій ітерації завдання (2019) найкращі результати становили: F1 = 0.829 для підзадачі A (детекція образливості), F1 = 0.755 для підзадачі B (таргетована або нетаргетована образа) та F1 = 0.660 для підзадачі C (ідентифікація мішені IND/GRP/OTH). Підходи на основі BERT переважали серед лідерів [8]. Пояснювальна компонента була розвинена в SemEval 2021 Task 5 Toxic Spans, де метою стало виокремлення токсичних фрагментів на рівні символів. Найкраща команда досягла character F1 = 70.83 %, що підтвердило практичну реалізованість спан-рівневого раціоналізування рішень трансформерами. Водночас зафіксовано помітну варіативність якості: окремі системи на основі BiLSTM CRF або ToxicBERT демонстрували F1 на рівні 62.23 %, що вказує на складність спан-детекції та чутливість до архітектури і налаштувань [9].

Дослідження 2023 і 2024 років засвідчили конкурентоспроможність донавчання сучасних трансформерів у мовно і ресурсно обмежених сценаріях. Для бенгальської мови повідомлялося про F1 = 0.87 із використанням Bangla BERT або Multilingual BERT [10]. У домені мікроблогів окремі інженерні рішення на локальних твіттер-корпусах досягали F1 = 0.91 за специфічних експериментальних умов, що підтверджує важливість доменної адаптації [11]. Огляди узагальнюють стабільну перевагу sentence або cross encoder трансформерів над традиційними моделями і підкреслюють значення урахування сеансового контексту для підвищення надійності детекції [12].

Для пояснюваних рішень широке застосування отримав датасет HateXplain, який поєднує клас мови ворожнечі, ціль висловлювання та раціоналі. Це створює умови для спільного навчання детекції і інтерпретації (мішень і фрагмент) та підвищує довіру до модераційних систем [13, 14].

Підсумовуючи, наявні результати окреслюють зрілість трансформерних підходів у бінарній і багаторівневій класифікації, а також у раціоналізації на рівні спанів. Недостатньо опрацьованими залишаються аспекти переходу від класифікації і пояснення до відновлення повної рольової структури висловлювань та надійної ідентифікації суб'єктів впливу, а також питання відтворюваності і масштабування таких рішень у керованому хмарному середовищі. Саме ці елементи визначають подальший вектор дослідження.

**Мета статті.** Сформувати та обґрунтувати об’єктно-орієнтований підхід до нейромережевого виявлення суб’єктів кібербулінгу за повідомленнями, що поєднує первинне визначення наявності кібербулінгу із подальшою синтаксико-семантичною інтерпретацією і відновленням рольових зв’язків «суб’єкт-дія-об’єкт». Передбачено розроблення об’єктної моделі предметної області, опис алгоритмічних етапів обробки, реалізацію прототипу у керованому хмарному середовищі та експериментальне оцінювання якості і продуктивності з акцентом на відтворюваність і масштабованість.

**Виклад основного матеріалу дослідження.** Запропонований об’єктно-орієнтований підхід до нейромережевого виявлення суб’єктів кібербулінгу за повідомленнями реалізує послідовну обробку текстових даних у керованому хмарному середовищі (рисунок 1). На вхід подаються неструктуровані повідомлення, які розглядаються як потенційні носії агресивної комунікації. Перший етап передбачає нейромережеве визначення наявності кібербулінгу: модель класифікує висловлювання за ознаками образливості, цькування чи вербального тиску, що дозволяє виокремити лише ті фрагменти, які потребують подальшої інтерпретації.

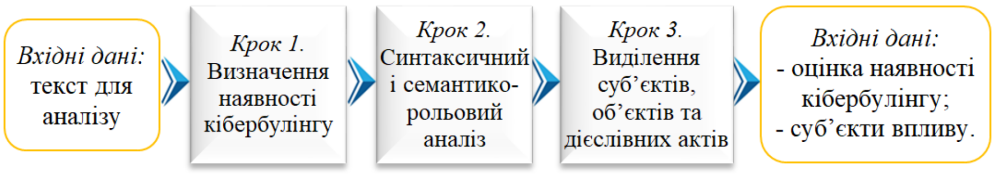


Рис. 1. Схема об’єктно-орієнтованого підходу до нейромережевого виявлення суб’єктів кібербулінгу

На наступному етапі здійснюється синтаксико-семантичний розбір повідомлень з реконструкцією граматичної структури та визначенням предикатів і лексичних носіїв дії. У межах об’єктно-орієнтованого подання домену відповідні сутності повідомлення, речення, предикат і рольова трійка використовуються для впорядкування результатів аналізу та їх подальшої інтерпретації.

Завершальний етап спрямований на встановлення суб’єктно-об’єктних відносин у висловлюваннях: визначається, хто ініціює мовленнєвий акт, на кого він спрямований і яким чином реалізується вплив.

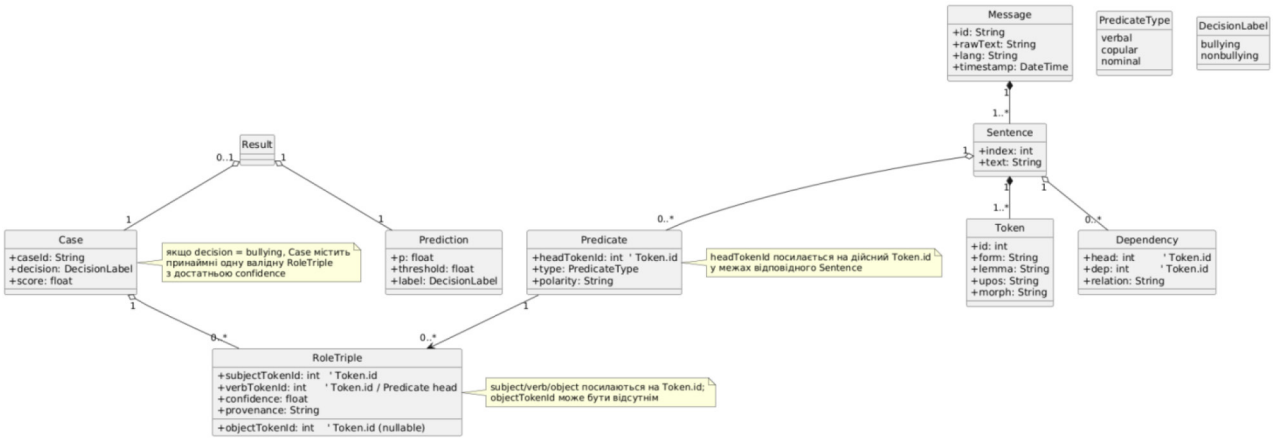


Рис. 2. UML-діаграма класів системи виявлення суб’єктів кібербулінгу за повідомленнями

Подана UML-модель (рисунок 2) формалізує об’єктно-орієнтоване подання задачі нейромережевого виявлення суб’єктів кібербулінгу у повідомленнях та відображає фактичну логіку реалізації. Вхідне повідомлення репрезентується класом «Message» із базовими метаданими; воно композиційно розчленовується на речення, кожне з яких містить токени й мережу залежностей, що задає синтаксичну структуру висловлювання. На цій структурі визначається предикат як ядро дії з фіксацією типу (дієслівний, копулярний або номінальний) і полярності. Така стратифікація рівнів, від тексту через речення і залежності до предиката забезпечує узгоджений перехід від поверхневої форми до змістової інтерпретації.

Рольова реконструкція реалізується через клас «RoleTriple», який фіксує трійку «суб’єкт, дія, об’єкт» з посиланнями на відповідні токени, а також зберігає впевненість і походження отриманих зв’язків. За наявності кількох актів мовленнєвого впливу трійки агрегуються у «Case», що містить підсумкове рішення та інтегральний бал і тим самим надає зручну одиницю для подальшого аналізу адресності. Первинне визначення факту кібербулінгу здійснюється нейромережевим класифікатором, результат якого акумулюється у класі «Prediction» у вигляді ймовірнісної оцінки, порогового значення та підсумкової мітки. Сукупний вихід конвеєра представлено класом «Result», який об’єднує рішення детектора з рольовою інтерпретацією у вигляді окремих трійок або цілісного випадку.

Запропонована об’єктна схема узгоджує дані і результати на всіх етапах обробки, дозволяє явно фіксувати посилання між рівнями представлення (від токенів до ролей) та забезпечує відтворюваність експериментів у керованому хмарному середовищі. Чітке розмежування сутностей і їхніх відповідальностей полегшує експериментальну валідацію, аудиту рішень і масштабування обробки, а також створює основу для подальшого оцінювання якості та продуктивності системи у прикладних сценаріях модерації.

Отримане структуроване подання випадку кібербулінгу поєднує факт агресії з чітко окресленою рольовою конфігурацією, що створює підґрунтя для аналізу адресності, динаміки взаємодій та подальшого практичного використання у керованому хмарному середовищі.

Запропонована архітектура виконання (рисунок 3) відображає розгортання підходу у керованому хмарному середовищі та структурована навколо єдиного Python-середовища, у межах якого функціонують основні компоненти обробки. Користувацький інтерфейс («UI/Demo») забезпечує подання вхідних повідомлень і перегляд результатів. Керування послідовністю етапів здійснює координатор обробки («Pipeline Coordinator»), який приймає повідомлення, ініціює первинне нейромережеве визначення наявності кібербулінгу, передає релевантні тексти на синтаксико-семантичний аналіз та збирає результати в узгодженому форматі.

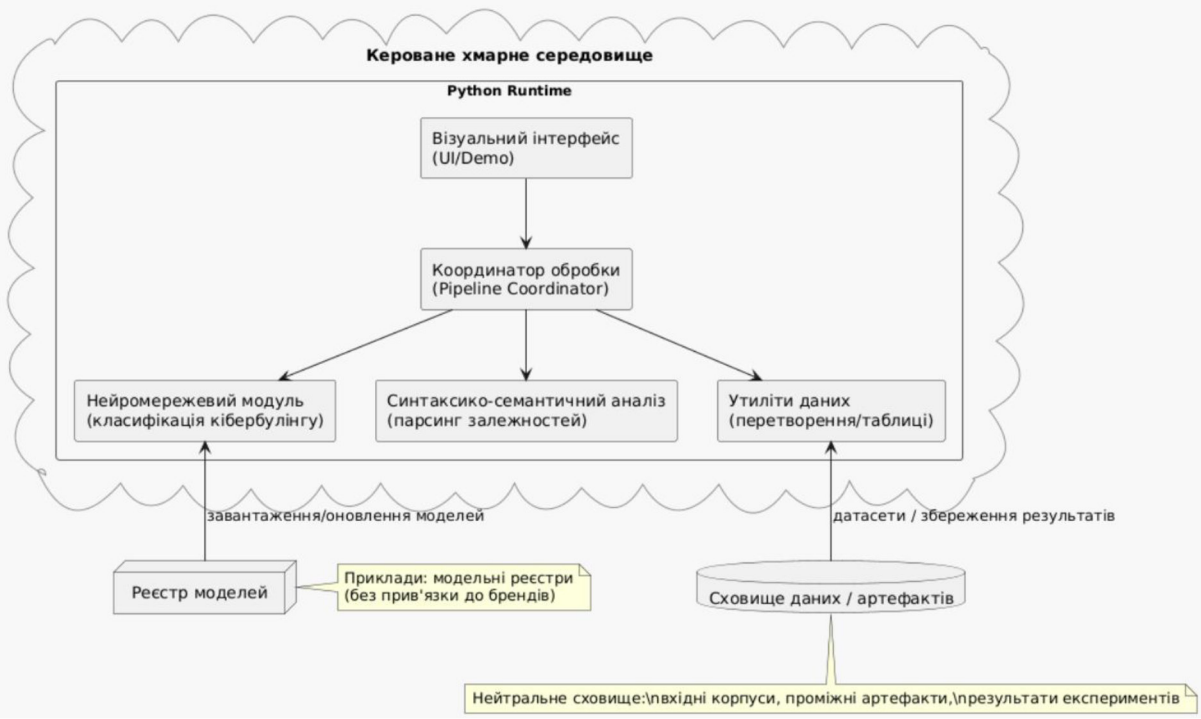


Рис. 3. Архітектура виконання у керованому хмарному середовищі

Нейромережевий модуль відповідає за класифікацію повідомлень щодо ознак кібербулінгу, повертаючи імовірнісне рішення, яке використовується як фільтр для подальшої інтерпретації. Блок синтаксико-семантичного аналізу виконує парсинг залежностей і формує підґрунтя для рольової реконструкції «суб’єкт-дія-об’єкт». Допоміжні утиліти даних здійснюють перетворення корпусів, формування підвибірок, кешування проміжних артефактів і підготовку підсумкових таблиць метрик, що спрощує відтворення експериментів.

Середовище взаємодіє з реєстром моделей, звідки здійснюється завантаження або оновлення модельних ваг та токенізаторів, і зі сховищем даних/артефактів, де розміщуються вхідні датасети, проміжні результати та підсумкові звіти. Така організація виключає прив’язку до конкретних сервісів, забезпечує контрольованість версій, можливість масштабування обчислень і прозоре трасування всіх кроків обробки від подання повідомлення до формування рішення про наявність кібербулінгу та отримання ролей учасників.

У роботі використано два відкриті корпуси: «Cyberbullying Classification» [15] та «Cyberbullying Detection» [16].

Корпус «Cyberbullying Classification» містить понад 47 тис. англомовних твітів із розміткою за шістьма категоріями: вікова, етнічна, гендерна, релігійна належність, інші прояви агресії та «не кібербулінг». Розподіл класів близький до збалансованого, що робить набір зручним для навчання і валідації бінарного детектора (об’єднання всіх проявів агресії в позитивний клас). Корпус відображає природну мову мікроблогів (сленг, орфографічні варіації, короткі висловлювання), однак не містить розмітки суб’єктів і об’єктів, тож не придатний для кількісної валідації рольових трійок.

Набір об'єднує тексти з різних платформ і використовується для двокласової постановки «кібербулінг / не кібербулінг». Різнорідність джерел підвищує узагальнюючу здатність моделі в задачі первинної детекції, проте наявний дисбаланс класів і відсутня деталізація типів агресії чи її адресності. У цій роботі корпус застосовано як допоміжний для перевірки узагальнюваності детектора; кількісна оцінка рольової інтерпретації на ньому не проводилась через брак відповідної розмітки.

Через відсутність у використаних корпусах «золотого стандарту» для рольових зв'язків (хто ініціатор, на кого спрямовано дію, яка дія), валідацію результатів рольової інтерпретації здійснювали два галузеві експерти та GPT-5 [17] як допоміжний суддя. Такий підхід забезпечує поєднання фахової оцінки з відтворюваною автоматизованою перевіркою.

Експерти незалежно анотували для кожного тестового прикладу трійку суб'єкт-дія-об'єкт і робили вердикт щодо коректності відновлення ролей системою. Узгодженість фіксували показниками міжекспертної згоди (відсоток повної згоди). У випадках розбіжностей застосовували схему більшості: два з трьох суддів формували остаточне рішення. GPT-5 використовували як третю сторону з чітким інструктажем, що вимагав: (1) вказати знайдені ролі, (2) навести опорні фрагменти тексту, (3) пояснити причину відхилення, якщо система помилилась. Вихід GPT-5 мав дорадчий характер і не замінював людське рішення.

У відкритих наборах даних, використаних для навчання детектора, відсутня рольова розмітка, тому потрібний людський еталон для перевірки відновлення суб'єктів. В свою чергу, залучення GPT-5 як додаткового судді підвищує відтворюваність і допомагає виявляти пропуски або неоднозначності, що залишилися поза увагою експертів. Також, така трирівнева схема скорочує витрати на повну ручну розмітку, зберігаючи при цьому якість і прозорість: фінальний вердикт завжди формується або повною згодою експертів, або більшістю з трьох суддів із фіксацією обґрунтувань.

У підсумку, обрана стратегія оцінювання дозволяє коректно перевіряти саме якість відновлення суб'єктів/об'єктів і дій, не обмежуючись лише метриками детекції кібербулінгу, та водночас відповідає вимогам відтворюваності, заявленим у меті дослідження.

Нейромережевий модуль первинної детекції реалізовано на основі трансформерного енкодера типу BERT [17] із донавчанням під двокласову постановку «кібербулінг» / «не кібербулінг». Передобробка обмежувалася стандартною токенізацією; усі прояви агресивної комунікації було зведено до позитивного класу. Оцінювання виконували на валідаційному піднаборі зі стратифікованим поділом даних. Запуски проводили у керованому хмарному середовищі з GPU класу T4, із фіксацією випадкових зерен і версій модельних ваг, що забезпечує відтворюваність без прив'язки до конкретної платформи.

Рольова інтерпретація ґрунтувалася на залежнісному аналізі речень і реконструкції трійок «суб'єкт-дія-об'єкт». За відсутності еталонної рольової розмітки кількісні метрики для цього етапу доповнювали контрактними перевірками узгодженості (валідність посилань на токени, єдиність предиката в реченні, несуперечливість ролей) та експертним рецензуванням, описаним у відповідному підрозділі. Така схема мінімізує залежність від конкретних реалізацій і водночас дає змогу коректно інтерпретувати отримані результати.

Нейромережевий модуль первинної детекції продемонстрував $F_1 = 0.98$ на валідаційних даних у двокласовій постановці «кібербулінг» / «не кібербулінг».

Через відсутність у відкритих корпусах «золотого стандарту» для ролей якість відновлення суб'єктів/об'єктів і предикатів оцінювали на експертно перевіреному піднаборі. У таблиці 1 наведено узагальнені показники (відносно узгоджених міток більшості).

Таблиця 1

**Метрики для виявлення суб'єктів кібербулінгу**

| Завдання | Показник |
|---|---|
| Визначення суб'єкта | Precision / Recall / $F_1$: 0.88 / 0.86 / 0.87 |
| Визначення об'єкта | Precision / Recall / $F_1$: 0.85 / 0.83 / 0.84 |
| Визначення дієслівного центру | Precision / Recall / $F_1$: 0.91 / 0.89 / 0.90 |
| Точне відновлення трійки | $F_1$: 0.76 |

Перед узгодженням рішень зафіксовано Cohen's $\kappa = 0.82$ при повній згоді 87 %. Це свідчить про високу відтворюваність критеріїв анотування і достатню надійність отриманих оцінок. Використання GPT-5 як третього судді мало допоміжний характер: модель надавала пояснення й опорні фрагменти, що спрощувало вирішення спірних випадків і підвищувало прозорість процедури; фінальні мітки завжди визначалися правилом більшості.

Спеціалізований конвеєр забезпечує детерміноване й відтворюване відновлення рольових зв'язків «суб'єкт-дія-об'єкт» з фіксованими порогами та контрактними перевірками, тож результати є аудитованими і придатними для регламентного використання. Натомість вихід GPT-5 суттєво залежить від версії та формулювання підказки, що ускладнює формальну верифікацію та дотримання нормативних вимог.

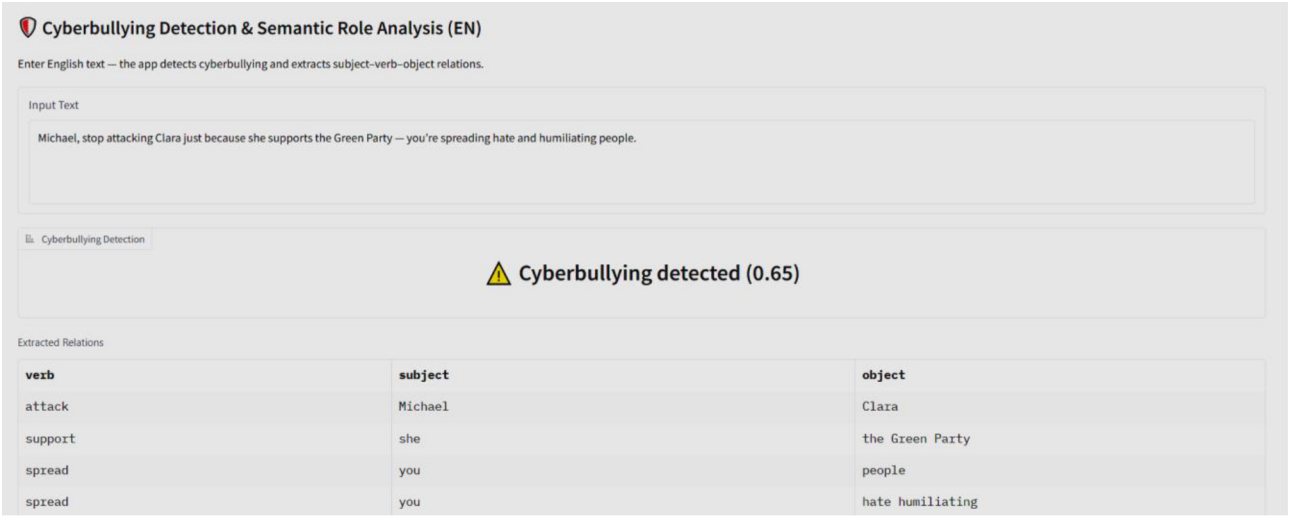


Рис. 4. Приклад роботи розробленого об'єктно-орієнтованого застосунку

Приклад роботи розробленого програмного забезпечення наведено на рисунку 4.

Запропонований об'єктно-орієнтований підхід забезпечує узгоджену послідовність обробки: від нейромережевої детекції кібербулінгу до синтаксико-семантичної реконструкції зв'язків «суб'єкт-дія-об'єкт», що реалізовано у керованому хмарному середовищі з фіксацією параметрів для відтворюваності. Архітектура та процедура оцінювання створюють передумови для масштабованого застосування у модераційних сценаріях і подальшого розширення на корпуси з детальнішою анотацією ролей.

**Висновки.** В межах проведеного дослідження було досягнуто поставлену мету: сформовано та обґрунтовано об'єктно-орієнтований підхід, у якому нейромережеве виявлення кібербулінгу поєднано з подальшою синтаксико-семантичною інтерпретацією й відновленням рольових зв'язків «суб'єкт-дія-об'єкт», з реалізацією у керованому хмарному середовищі. У межах двокласової постановки первинний модуль на основі BERT продемонстрував $F_1$ на рівні 0.98, що підтверджує достатній рівень фільтрації повідомлень перед рольовим аналізом. На експертно-перевіреному піднаборі відтворення ролей для визначення суб'єкта зафіксовано значення 0.88, 0.86, 0.87 за метриками Precision, Recall, $F_1$ відповідно, для об'єкта 0.85, 0.83, 0.84, а для дієслівного центру 0.91, 0.89, 0.90, тоді як точне відновлення трійки дало $F_1$ у розмірі 0.76. Міжекспертна узгодженість становила к Коена 0.82 при 87 % повної згоди, що підтверджує надійність еталонних міток і коректність процедури оцінювання. Сукупно результати свідчать, що запропонована архітектура забезпечує відтворюваність, аудитованість і практичну придатність: вона не лише фіксує факт агресивної комунікації, а й надає структуровану інформацію про адресність мовленнєвого впливу, створюючи підґрунтя для впровадження в модераційні системи та подальшого розширення на корпуси з детальною рольовою розміткою.

**Список використаних джерел:**

1. A Human-Centered Systematic Literature Review of Cyberbullying Detection Algorithms / S. Kim та ін. *Proceedings of the ACM on Human-Computer Interaction*. 2021. Т. 5, CSCW2. С. 1–34. https://doi.org/10.1145/3476066 (дата звернення: 07.11.2025).

2. Paul S., Saha S., Hasanuzzaman M. Identification of cyberbullying: A deep learning based multimodal approach. *Multimedia Tools and Applications*. 2020. https://doi.org/10.1007/s11042-020-09631-w (дата звернення: 07.11.2025).

3. Human Activity Recognition for the Identification of Bullying and Cyberbullying Using Smartphone Sensors / V. Gattulli et al. *Electronics*. 2023. Vol. 12, no. 2. P. 261. https://doi.org/10.3390/electronics12020261 (дата звернення: 07.11.2025).

4. Method for neural network cyberbullying detection in text content with visual analytic / I. Krak et al. *CEUR Workshop Proceedings*. 2025. Vol. 3917, PP. 298–309. URL: https://ceur-ws.org/Vol-3917/paper57.pdf (дата звернення: 07.11.2025).

5. Method for cyberbullying neuronetwork detection using cloud services and object-oriented model / M. Молчанова та ін. *Herald of Khmelnytskyi National University. Technical sciences*. 2024. Vol. 333, no. 2. P. 200–206. https://doi.org/10.31891/2307-5732-2024-333-2 (дата звернення: 07.11.2025).

6. Verma R., Kumar K., Verma H. K. Code smell prioritization in object-oriented software systems: A systematic literature review. *Journal of Software: Evolution and Process*. 2023. https://doi.org/10.1002/smr.2536 (дата звернення: 07.11.2025).

7. Load Balancing in cloud Environment: A State of-the-Art Review / Y. Lohumi et al. *IEEE Access*. 2023. P. 1. https://doi.org/10.1109/access.2023.3337146 (дата звернення: 07.11.2025).

8. OffensEval 2023: Offensive language identification in the age of Large Language Models / M. Zampieri et al. *Natural Language Engineering*. 2023. Vol. 29, no. 6. P. 1416–1435. https://doi.org/10.1017/s1351324923000517 (дата звернення: 07.11.2025).

9. SemEval-2021 Task 5: Toxic Spans Detection / J. Pavlopoulos et al. *Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021)*, Online. Stroudsburg, PA, USA, 2021. https://doi.org/10.18653/v1/2021.semeval-1.6 (дата звернення: 07.11.2025).

10. 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. Article No. 100104. https://doi.org/10.1016/j.nlp.2024.100104 (дата звернення: 07.11.2025).

11. Aliyeva Ç. O., Yağanoğlu M. Deep learning approach to detect cyberbullying on twitter. *Multimedia Tools and Applications*. 2024. https://doi.org/10.1007/s11042-024-19869-3 (дата звернення: 07.11.2025).

12. Yi P., Zubiaga A. Session-based cyberbullying detection in social media: A survey. *Online Social Networks and Media*. 2023. Vol. 36. P. 100250. https://doi.org/10.1016/j.osnem.2023.100250 (дата звернення: 07.11.2025).

13. HateXplain: A Benchmark Dataset for Explainable Hate Speech Detection / B. Mathew et al. *Proceedings of the AAAI Conference on Artificial Intelligence*. 2021. Vol. 35, no. 17. P. 14867–14875. https://doi.org/10.1609/aaai.v35i17.17745 (дата звернення: 07.11.2025).

14. Hate-speech-CNERG/hatexplain · Datasets at Hugging Face. *Hugging Face – The AI community building the future*. URL: https://huggingface.co/datasets/Hate-speech-CNERG/hatexplain (дата звернення: 07.11.2025).

15. Cyberbullying Classification. *Kaggle*. URL: https://www.kaggle.com/datasets/andrewmvd/cyberbullying-classification (дата звернення: 07.11.2025).

16. Cyberbullying Detection. *Kaggle*. URL: https://www.kaggle.com/datasets/gbiamgaurav/cyberbullying-detection (дата звернення: 07.11.2025).

17. GPT-5. OpenAI. URL: https://openai.com/gpt-5/ (дата звернення: 07.11.2025).

18. BERT applications in natural language processing: a review / N. M. Gardazi et al. *Artificial Intelligence Review*. 2025. Vol. 58, no. 6. https://doi.org/10.1007/s10462-025-11162-5 (дата звернення: 07.11.2025).

**References:**

1. Kim, S., et al. (2021). A human-centered systematic literature review of cyberbullying detection algorithms. Proceedings of the ACM on Human-Computer Interaction, 5(CSCW2), 1–34. https://doi.org/10.1145/3476066

2. Paul, S., Saha, S., & Hasanuzzaman, M. (2020). Identification of cyberbullying: A deep learning-based multimodal approach. *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-020-09631-w

3. Gattulli, V., et al. (2023). Human activity recognition for the identification of bullying and cyberbullying using smartphone sensors. *Electronics*, 12(2), 261. https://doi.org/10.3390/electronics12020261

4. Krak I., et al. (2025). Method for neural network cyberbullying detection in text content with visual analytic. *CEUR Workshop Proceedings*, 2025, vol. 3917, 298–309. Retrieved from: https://ceur-ws.org/Vol-3917/paper57.pdf

5. Molchanova, M., et al. (2024). Method for cyberbullying neuronetwork detection using cloud services and object-oriented model. *Herald of Khmelnytskyi National University: Technical Sciences*, 333(2), 200–206. https://doi.org/10.31891/2307-5732-2024-333-2

6. Verma, R., Kumar, K., & Verma, H. K. (2023). Code smell prioritization in object-oriented software systems: A systematic literature review. *Journal of Software: Evolution and Process*. https://doi.org/10.1002/smr.2536

7. Lohumi, Y., et al. (2023). Load balancing in cloud environment: A state-of-the-art review. *IEEE Access*, 1. https://doi.org/10.1109/access.2023.3337146

8. Zampieri, M., et al. (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

9. Pavlopoulos, J., et al. (2021). SemEval-2021 Task 5: Toxic spans detection. In Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021). Stroudsburg, PA: Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.semeval-1.6

10. 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

11. Aliyeva, Ç. O., & Yağanoğlu, M. (2024). Deep learning approach to detect cyberbullying on Twitter. *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-024-19869-3

12. Yi, P., & Zubiaga, A. (2023). Session-based cyberbullying detection in social media: A survey. *Online Social Networks and Media*, 36, 100250. https://doi.org/10.1016/j.osnem.2023.100250

13. Mathew, B., et al. (2021). HateXplain: A benchmark dataset for explainable hate speech detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(17), 14867–14875. https://doi.org/10.1609/aaai.v35i17.17745

14. Hate-speech-CNERG/hatexplain. (n.d.). Datasets at Hugging Face – The AI community building the future. Hugging Face. Retrieved from: https://huggingface.co/datasets/Hate-speech-CNERG/hatexplain (Accessed November 7, 2025)

15. Cyberbullying Classification. (n.d.). *Kaggle*. Retrieved from: https://www.kaggle.com/datasets/andrewmvd/cyberbullying-classification (Accessed November 7, 2025)

16. Cyberbullying Detection. (n.d.). *Kaggle*. Retrieved from: https://www.kaggle.com/datasets/gbiamgaurav/cyberbullying-detection (Accessed November 7, 2025)

17. OpenAI. (2025). *GPT-5*. Retrieved from: https://openai.com/gpt-5/ (Accessed November 7, 2025)

18. Gardazi, N. M., et al. (2025). BERT applications in natural language processing: A review. *Artificial Intelligence Review*, 58(6). https://doi.org/10.1007/s10462-025-11162-5

Дата першого надходження статті до видання: 09.11.2025
Дата прийняття статті до друку після рецензування: 10.12.2025
Дата публікації (оприлюднення) статті 27.01.2026