МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ПОЛІСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

Факультет інформаційних технологій,
обліку та фінансів
Кафедра комп'ютерних технологій
і моделювання систем

Кваліфікаційна робота
на правах рукопису

Ломберг Едуард Вікторович
(прізвище, ім'я, по батькові здобувача освіти)

УДК 004.8:004.056.53:004.738.5

КВАЛІФІКАЦІЙНА РОБОТА

Інтелектуальна система виявлення ботів у соціальних мережах на основі машинного навчання та аналізу аномальної поведінки користувачів
(тема роботи)

122 "Комп'ютерні науки"
(шифр і назва спеціальності)

Подається на здобуття освітнього ступеня бакалавр

кваліфікаційна робота містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело

_____
(підпис, ініціали та прізвище здобувача освіти)

Керівник роботи
Ковальчук М. О.
(прізвище, ім'я, по батькові)

к.п.н. ,доцент
(науковий ступінь, вчене звання)

Житомир – 2026

Висновок кафедри комп'ютерних технологій і моделювання систем
за результатами попереднього захисту: допущений_

Протокол засідання кафедри комп'ютерних технологій і моделювання систем
№ 20 від «05» червня 2026 р.
Завідувач кафедри _____
к. пед. н., доцент _____ КОВАЛЬЧУК М. О.
(науковий ступінь, вчене звання) (підпис) (прізвище, ім'я, по батькові)
«____» _____ 20____ р.

Результати захисту кваліфікаційної роботи

Здобувач вищої освіти _____ захистив (ла)
(прізвище ,ім'я, по батькові)
кваліфікаційну роботу з оцінкою:

сума балів за 100-бальною шкалою _____
за шкалою ECTS _____
за національною шкалою _____

Секретар ЕК
лаборант кафедри _____ КОВАЛЬЧУК В.В.
(науковий ступінь, вчене звання) (підпис) (прізвище, ім'я, по батькові)

АНОТАЦІЯ

Ломберг Е.В. Інтелектуальна система виявлення ботів у соціальних мережах на основі машинного навчання та аналізу аномальної поведінки користувачів. – Кваліфікаційна робота на здобуття освітнього ступеня бакалавра за спеціальністю 122 «Комп'ютерні науки». – Поліський національний університет, Житомир, 2026.

У роботі розглянуто проблему виявлення ботоподібної активності на платформі YouTube. Проведено аналіз існуючих методів виявлення ботоподібної активності, спроєктовано та реалізовано інтелектуальну систему аналізу поведінки користувачів.

Розроблена система використовує дані, отримані через YouTube Data API v3, формує поведінкові ознаки користувачів, виконує пошук аномалій за допомогою алгоритму Isolation Forest та виявляє групи схожих користувачів із використанням алгоритму DBSCAN. Результатом роботи системи є оцінка ризику ботоподібної активності та пояснення виявлених аномалій.

Практичним результатом роботи є веборієнтована система автоматизованого аналізу активності користувачів платформи YouTube.

Ключові слова: YouTube, автоматизований обліковий запис, машинне навчання, виявлення аномалій, поведінкові ознаки, метод ізоляційного лісу, DBSCAN.

SUMMARY

Lomberg E.V. Intelligent Bot Detection System in Social Networks Based on Machine Learning and Anomalous User Behavior Analysis. – Bachelor's Qualification Thesis in Specialty 122 "Computer Science". – Polissia National University, Zhytomyr, 2026.

The thesis addresses the problem of detecting bot-like activity on the YouTube platform. Existing bot detection approaches were analyzed, and an intelligent user behavior analysis system was designed and implemented.

The developed system uses data obtained through YouTube Data API v3, extracts behavioral features, detects anomalies using the Isolation Forest algorithm,

and identifies groups of similar users using the DBSCAN algorithm. The system provides bot activity risk assessment and explanations of detected anomalies.

The practical result of the work is a web-based system for automated analysis of YouTube user activity.

Keywords: YouTube, bot detection, machine learning, anomaly detection, behavioral features, Isolation Forest, DBSCAN.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

API – Application Programming Interface (інтерфейс програмування застосунків)

DBSCAN – Density-Based Spatial Clustering of Applications with Noise (алгоритм кластеризації на основі щільності)

HTML – HyperText Markup Language (мова розмітки гіпертекстових документів)

HTTP – HyperText Transfer Protocol (протокол передавання гіпертексту)

IDEF0 – Integration Definition for Function Modeling (методологія функціонального моделювання)

JSON – JavaScript Object Notation (текстовий формат обміну даними)

ML – Machine Learning (машинне навчання)

ORM – Object Relational Mapping (технологія об'єктно-реляційного відображення)

PostgreSQL – система керування реляційними базами даних

React – JavaScript-бібліотека для розробки користувацьких інтерфейсів

REST – Representational State Transfer (архітектурний стиль взаємодії вебсервісів)

URL – Uniform Resource Locator (уніфікований локатор ресурсу)

UML – Unified Modeling Language (уніфікована мова моделювання)

YouTube Data API v3 – програмний інтерфейс доступу до даних платформи YouTube

БД – база даних

ІС – інформаційна система

ВСТУП

У сучасному інформаційному середовищі YouTube є однією з провідних платформ для поширення інформації, комунікації та формування суспільної думки. Водночас зі зростанням популярності платформи збільшується кількість автоматизованих облікових записів, які можуть використовуватися для поширення спаму, маніпулювання громадською думкою та штучного підвищення активності навколо відеоконтенту.

Сучасні бот-мережі здатні імітувати поведінку реальних користувачів, що ускладнює їх виявлення традиційними методами, заснованими на фіксованих правилах або ручному аналізі. Тому актуальним є використання методів машинного навчання та аналізу аномальної поведінки, які дозволяють автоматизувати процес виявлення підозрілих облікових записів.

Метою кваліфікаційної роботи є розробка інтелектуальної системи виявлення ботоподібної активності на платформі YouTube на основі машинного навчання та аналізу аномальної поведінки користувачів.

Для досягнення поставленої мети необхідно вирішити такі завдання:

- провести аналіз предметної області та існуючих методів виявлення ботоподібної активності;
- дослідити особливості використання YouTube Data API v3 для отримання даних;
- сформувати функціональні та нефункціональні вимоги до системи;
- спроєктувати архітектуру системи та структуру бази даних;
- реалізувати збір і попередню обробку даних YouTube;
- розробити механізм формування поведінкових ознак користувачів;
- реалізувати алгоритми виявлення аномальної поведінки та оцінювання ризику;

Об’єктом дослідження є процес аналізу поведінки користувачів платформи YouTube з метою виявлення ознак автоматизованої активності.

Предметом дослідження є методи машинного навчання, аналізу аномальної поведінки та формування поведінкових ознак користувачів для виявлення ботоподібної активності на YouTube.

Під час виконання роботи використано методи аналізу наукових джерел (для дослідження актуальності проблеми поширення бот-мереж та оцінювання існуючих аналогів), системного аналізу (для формування вимог до системи й визначення структури вхідних і вихідних даних), UML та IDEF0-моделювання (для проєктування функціональної архітектури, взаємодії компонентів та структури бази даних), машинного навчання (для автоматизації оцінювання ризику ботоподібної активності на основі сформованих ознак), аналізу аномалій (для виявлення нетипових профілів поведінки користувачів за допомогою алгоритму Isolation Forest), кластерного аналізу (для групування схожих акаунтів та пошуку координованих груп ботів за допомогою алгоритму DBSCAN) та програмної інженерії (для програмної реалізації, тестування та розгортання веборієнтованої інтелектуальної системи).

Практичне значення отриманих результатів полягає у створенні програмної системи, яка забезпечує автоматизований аналіз активності користувачів YouTube, виявлення аномальної поведінки та оцінювання ризику ботоподібної активності.

За результатами дослідження опубліковано дві наукові праці, у яких висвітлено основні результати виконаної роботи:

1. Ломберг Е. В. Виявлення координованої активності ботів у коментарях YouTube на основі аналізу текстових, часових і поведінкових ознак: **збірник праць учасників Всеукраїнської науково-практичної конференції здобувачів вищої освіти і молодих вчених**, м. Житомир, 22 травня 2026 р. Житомир : Поліський національний університет, 2026. С. 59–60.

2. Ломберг Е. В. Вплив соціальних ботів на сучасний інформаційний простір. **New Horizons in Scientific Research: Challenges and Solutions** : матеріали наукової конференції, 15–17 червня 2026 р С. 215–216.

Кваліфікаційна робота складається зі вступу, трьох розділів, висновків, списку використаних джерел та додатків. У першому розділі виконано аналіз предметної області та сформовано вимоги до системи. У другому розділі розроблено архітектуру, структуру даних та алгоритмічне забезпечення системи. У третьому розділі наведено опис реалізації, результати тестування та приклади використання розробленої системи.

РОЗДІЛ 1 ТЕОРЕТИЧНИЙ АНАЛІЗ ОСОБЛИВОСТЕЙ ПРЕДМЕТНОЇ ОБЛАСТІ ДОСЛІДЖЕННЯ

1.1 Аналіз предметної області виявлення ботоподібної активності на платформі YouTube

У сучасному цифровому середовищі соціальні мережі та відеохостинги стали одним із основних джерел отримання інформації, комунікації та формування суспільної думки. Однією з найбільших платформ для розміщення та поширення відеоконтенту є YouTube, яка об'єднує мільйони користувачів по всьому світу та забезпечує широкий спектр можливостей для взаємодії через перегляди, коментарі, вподобання та підписки.

Зі зростанням популярності платформи збільшується і кількість автоматизованих облікових записів, відомих як боти. Такі акаунти можуть використовуватися для накрутки переглядів, штучного збільшення кількості підписників, поширення спаму, генерації коментарів та маніпулювання популярністю відеоконтенту. У результаті цього спотворюються статистичні показники активності, знижується достовірність аналітичних даних та створюються передумови для інформаційних маніпуляцій[2].

Однією з основних проблем сучасних платформ є складність виявлення автоматизованих облікових записів. Сучасні бот-мережі здатні імітувати поведінку реальних користувачів, використовувати різні часові інтервали між діями та генерувати контент, схожий на повідомлення звичайних користувачів. У зв'язку з цим традиційні методи виявлення, що базуються на ручному аналізі або наборі фіксованих правил, поступово втрачають свою ефективність.

Актуальність проблеми підтверджується результатами сучасних досліджень. Так, у 2025 році було проведено аналіз понад 97 тисяч коментарів до більш ніж 2600 відео на платформі YouTube. За результатами дослідження близько 14 % коментарів у досліджуваній вибірці були створені автоматизованими обліковими записами, а майже 39 % відео зазнали впливу організованих бот-мереж. Також було виявлено понад 6 тисяч акаунтів із ознаками автоматизованої діяльності. Отримані результати свідчать про значне

поширення ботів на платформі та необхідність розробки ефективних засобів їх виявлення.[10]

Для вирішення задачі виявлення ботоподібної активності використовуються різноманітні підходи, серед яких аналіз поведінкових характеристик користувачів, виявлення аномальної активності та застосування алгоритмів машинного навчання. Джерелом інформації для такого аналізу можуть виступати дані про активність користувачів, отримані через YouTube Data API v3, зокрема інформація про коментарі, канали, частоту взаємодії з контентом та інші поведінкові показники[18].

Аналіз сучасного стану розробки методів протидії автоматизованій деструктивній діяльності в інтернет-середовищі свідчить, що більшість існуючих систем детекції орієнтована на текстоцентричні платформи (зокрема X/Twitter) або використовує редукований набір поведінкових характеристик. Подібний асиметричний фокус досліджень зумовлює гостру науково-практичну потребу у розробці спеціалізованої інтелектуальної системи, яка здатна здійснювати комплексний мультимодальний аналіз специфічних поведінкових ознак користувачів відеохостингу YouTube, ідентифікувати латентні поведінкові аномалії та виконувати прецизійну класифікацію облікових записів на основі адаптивних методів машинного навчання.

1.2 Аналіз існуючих методів та засобів виявлення ботоподібної активності

Для виявлення автоматизованих облікових записів у соціальних мережах застосовуються різні підходи, які відрізняються принципом роботи, точністю виявлення та складністю реалізації. Найбільш поширеними є методи, що базуються на використанні набору правил, аналізі поведінкових характеристик користувачів, алгоритмах машинного навчання та виявленні аномальної активності[20]. Кожен із зазначених підходів має власні переваги та обмеження, тому доцільно провести їх порівняльний аналіз. Порівняння основних методів виявлення ботоподібної активності наведено у таблиці 1.1.

Таблиця 1.1 – Порівняльна характеристика методів виявлення ботоподібної активності

Метод	**Принцип роботи**	**Переваги**	**Недоліки**
Rule-Based	Використання набору правил та порогових значень	Простота реалізації, швидкодія	Низька адаптивність до нових ботів
Поведінковий аналіз	Аналіз активності та шаблонів поведінки користувача	Виявлення прихованих закономірностей	Потребує значного обсягу даних
Машинне навчання	Класифікація акаунтів на основі ознак	Висока точність, адаптивність	Необхідність навчальних даних
Виявлення аномалій	Пошук нетипової поведінки користувачів	Може працювати без розмічених даних	Можливі хибні спрацьовування

Як видно з таблиці 1.1, використання набору правил є найпростішим способом виявлення ботоподібної активності, однак ефективність такого підходу суттєво знижується при появі нових типів автоматизованих акаунтів. Поведінковий аналіз дозволяє досліджувати особливості активності користувачів та виявляти приховані закономірності їхньої поведінки. Методи машинного навчання забезпечують високу точність класифікації та здатні адаптуватися до змін у поведінці ботів, проте потребують навчальних даних для побудови моделей. Методи виявлення аномалій дають можливість знаходити нетипову активність навіть без попередньо розмічених вибірок, що є важливою перевагою під час аналізу нових або маловідомих типів ботів.

Окрім окремих методів, існують спеціалізовані програмні рішення, призначені для виявлення ботоподібної активності та аналізу підозрілої активності у соціальних мережах. Такі системи використовують один або декілька підходів одночасно та забезпечують автоматизований аналіз великих обсягів даних. Для визначення можливості їх використання у задачі виявлення ботоподібної активності на платформі YouTube було проведено порівняння найбільш відомих рішень.

Серед найбільш відомих програмних рішень для виявлення ботоподібної активності та аналізу автоматизованої активності можна виділити Botometer, Bot

Sentinel, Hoaxy та BotHunter[27]. Дані системи використовують різні підходи до аналізу поведінки користувачів, поширення інформації та класифікації облікових записів. Для оцінки можливості їх застосування у задачі виявлення ботоподібної активності на платформі YouTube було проведено їх порівняльний аналіз.

Таблиця 1.2 – Порівняння існуючих систем виявлення ботоподібної активності.

Система	**Соціальна мережа**	**Основний підхід**	**Переваги**	**Недоліки**
Botometer	X (Twitter)	Машинне навчання	Висока точність оцінки акаунтів	Не підтримує YouTube
Bot Sentinel	X (Twitter)	Поведінковий аналіз	Аналіз токсичної та автоматизованої активності	Орієнтований на X
Hoaxy	Соціальні мережі	Аналіз поширення інформації	Візуалізація мережевих зв'язків	Не виконує повноцінну класифікацію ботів
BotHunter	Соціальні мережі	Машинне навчання та мережевий аналіз	Комплексний підхід	Обмежена підтримка платформ

Аналіз даних таблиці 1.2 показує, що більшість існуючих систем орієнтовані переважно на соціальну мережу X (Twitter) або інші платформи та не забезпечують повноцінної підтримки YouTube. Незважаючи на використання сучасних методів машинного навчання, поведінкового аналізу та дослідження мережевих зв'язків, розглянуті рішення мають обмеження щодо збору та аналізу поведінкових характеристик користувачів відеохостингу YouTube. Це ускладнює їх безпосереднє застосування для задачі, що розглядається у даній роботі.

Проведений аналіз методів і програмних засобів виявлення ботоподібної активності показав, що найбільш перспективним підходом є поєднання аналізу поведінкових характеристик користувачів, алгоритмів машинного навчання та методів виявлення аномалій. Саме такий підхід дозволяє забезпечити високу точність класифікації акаунтів та адаптивність до нових типів автоматизованої активності. Тому зазначені методи покладено в основу розроблюваної

інтелектуальної системи виявлення ботоподібної активності на платформі YouTube.

Для задачі виявлення ботоподібної активності на платформі YouTube особливий інтерес становить використання YouTube Data API v3, оскільки цей інтерфейс надає можливість легально отримувати публічно доступні дані про канали, відео та коментарі без застосування методів веб-скрапінгу[3]. На відміну від збору інформації шляхом аналізу HTML-сторінок, використання API забезпечує стабільний доступ до структурованих даних, зменшує ризик блокування за IP-адресою та не залежить від змін у структурі вебінтерфейсу або механізмів захисту, реалізованих за допомогою JavaScript. Водночас використання API має певні обмеження, пов’язані з квотами на кількість запитів та доступністю лише тих даних, які офіційно надаються платформою.

1.3 Формування вимог до інтелектуальної системи виявлення ботоподібної активності

Проведений аналіз предметної області та існуючих методів виявлення ботоподібної активності показав необхідність створення спеціалізованої системи, здатної автоматично виявляти підозрілу активність користувачів на платформі YouTube. Для забезпечення ефективної роботи майбутньої системи необхідно сформувати функціональні та нефункціональні вимоги, визначити основні вхідні та вихідні дані, а також встановити вимоги до обробки інформації та формування результатів аналізу.

До основних функціональних вимог розроблюваної системи належать збір даних про користувачів та їхню активність на платформі YouTube, формування набору поведінкових характеристик[22], виявлення аномальної активності, виявлення аномальної поведінки за допомогою алгоритмів машинного навчання та формування звіту про результати аналізу. Крім того, система повинна забезпечувати збереження результатів перевірок для їх подальшого використання та аналізу.

Таблиця 1.3 – Функціональні вимоги до інтелектуальної системи виявлення ботоподібної активності на платформі YouTube

Код	**Функціональна вимога**
F1	Прийом URL або ідентифікатора YouTube-каналу чи відео для аналізу
F2	Отримання даних через YouTube Data API v3
F3	Збір інформації про канали, відео, коментарі та авторів коментарів
F4	Формування набору поведінкових характеристик користувачів
F5	Виявлення аномальної поведінки користувачів
F6	Виявлення аномальної поведінки за допомогою алгоритмів машинного навчання
F7	Формування та відображення результатів аналізу
F8	Генерування звіту про результати перевірки
F9	Збереження результатів аналізу в базі даних
F10	Ведення журналу подій та статистики роботи системи

Як видно з таблиці 1.3, основні функції системи охоплюють повний цикл аналізу даних: від отримання інформації з платформи YouTube до формування підсумкового звіту для користувача. Особливістю розроблюваної системи є поєднання аналізу поведінкових характеристик користувачів, алгоритмів машинного навчання та методів виявлення аномалій, що дозволяє підвищити точність виявлення ботоподібної активності та забезпечити адаптацію до нових типів автоматизованої активності[21].

Окрім функціональних можливостей, важливим аспектом розробки системи є забезпечення її надійності, швидкодії, безпеки та зручності використання. З цією метою були сформовані нефункціональні вимоги, які визначають якісні характеристики майбутньої системи та умови її ефективної експлуатації.

Таблиця 1.4 – Нефункціональні вимоги до інтелектуальної системи виявлення ботоподібної активності на платформі YouTube

Код	**Нефункціональна вимога**
NF1	Забезпечення стабільної та безперервної роботи системи
NF2	Швидка обробка запитів користувачів та формування результатів аналізу
NF3	Можливість розширення системи новими алгоритмами аналізу та класифікації
NF4	Забезпечення зручного та зрозумілого інтерфейсу користувача
NF5	Раціональне використання квот YouTube Data API v3
NF6	Коректне збереження результатів аналізу та журналів роботи системи
NF7	Можливість роботи з великими обсягами даних без суттєвого зниження продуктивності

Як видно з таблиці 1.4, нефункціональні вимоги визначають якісні характеристики системи та забезпечують її ефективне використання на практиці. Особливу увагу приділено надійності роботи, швидкодії та раціональному використанню ресурсів YouTube Data API v3. Виконання зазначених вимог дозволить забезпечити стабільне функціонування системи та підвищити ефективність процесу виявлення ботоподібної активності.

Для забезпечення роботи інтелектуальної системи необхідно визначити основні вхідні та вихідні дані, які використовуються під час аналізу. Вхідними даними є інформація, що надходить до системи від користувача та з платформи YouTube через YouTube Data API v3. На основі цих даних формуються поведінкові характеристики користувачів, виконується пошук аномалій та класифікація облікових записів.

На основі отриманих вхідних даних система виконує аналіз активності користувачів, формує набір ознак для подальшої обробки алгоритмами машинного навчання та визначає наявність ознак автоматизованої поведінки.

На основі отриманої інформації формуються поведінкові характеристики користувачів, здійснюється виявлення аномальної активності та виконується класифікація облікових записів. Результатом роботи системи є набір вихідних

даних, що містить результати аналізу, виявлені аномалії та підсумковий висновок щодо належності акаунта до категорії ботів.

Таблиця 1.5 – Вхідні дані інтелектуальної системи виявлення ботоподібної активності на платформі YouTube

Код	**Вхідні дані**	**Призначення**
IN1	URL або ідентифікатор YouTube-відео	Отримання інформації про відео
IN2	Дані про канал	Аналіз характеристик каналу
IN3	Дані про відео	Аналіз показників відео
IN4	Коментарі користувачів	Аналіз текстової активності
IN5	Дані про авторів коментарів	Формування поведінкових характеристик
IN6	Статистичні показники активності	Виявлення аномальної поведінки
IN7	Поведінкові характеристики користувачів	Використання в алгоритмах машинного навчання

Як видно з таблиці 1.5, основними джерелами даних для роботи системи є відомості про канали, відео та активність користувачів, отримані через YouTube Data API v3. Отримана інформація використовується для формування поведінкових характеристик та подальшого аналізу із застосуванням методів машинного навчання й виявлення аномалій.

Висновки до розділу 1

У першому розділі було проведено аналіз предметної області виявлення ботоподібної активності на платформі YouTube та досліджено актуальність даної проблеми. Встановлено, що поширення автоматизованих облікових записів негативно впливає на достовірність статистичних показників, якість взаємодії користувачів та інформаційне середовище платформи в цілому.

Було розглянуто основні методи виявлення ботоподібної активності, серед яких використання набору правил, аналіз поведінкових характеристик користувачів, алгоритми машинного навчання та методи виявлення аномалій. Проведений аналіз показав, що найбільш перспективним підходом є поєднання аналізу поведінкових характеристик, алгоритмів машинного навчання та методів виявлення аномальної активності.

Також було виконано порівняльний аналіз існуючих програмних рішень для виявлення ботоподібної активності. Встановлено, що більшість розглянутих

систем орієнтовані на інші соціальні мережі та не забезпечують повноцінної підтримки платформи YouTube, що обмежує можливість їх використання для вирішення поставленої задачі.

На основі проведеного аналізу було сформовано функціональні та нефункціональні вимоги до розроблюваної інтелектуальної системи, визначено основні вхідні та вихідні дані, а також окреслено вимоги до процесу аналізу поведінкових характеристик користувачів. Отримані результати є основою для подальшого проектування архітектури системи, розробки моделей даних та побудови алгоритмів виявлення ботоподібної активності, що будуть розглянуті у наступному розділі.

РОЗДІЛ 2. ПРОЄКТУВАННЯ ІНТЕЛЕКТУАЛЬНОЇ СИСТЕМИ ВИЯВЛЕННЯ БОТОПОДІБНОЇ АКТИВНОСТІ

2.1 Моделювання функціональних вимог до системи

Після формування вимог до інтелектуальної системи виявлення ботоподібної активності на платформі YouTube необхідно виконати моделювання її функціональних можливостей. Використання UML-діаграм дозволяє формалізувати вимоги до системи, визначити взаємодію користувачів із програмним забезпеченням та описати основні сценарії виконання функцій.

Для моделювання функціональних вимог було використано діаграму прецедентів та діаграму діяльності. Дані моделі дозволяють відобразити взаємодію користувачів із системою та послідовність виконання процесів аналізу облікових записів.

Для формалізації функціональних вимог інтелектуальної системи виявлення ботоподібної активності в соцмережах використано UML-діаграму прецедентів. Діаграма відображає основних акторів системи, їх взаємодію з програмним забезпеченням та функції, необхідні для аналізу облікових записів і виявлення ботоподібної активності. Діаграма прецедентів наведена в додатку А.

Діаграма відображає основні сценарії взаємодії користувачів із системою та демонструє розподіл функцій між акторами. Вона наочно демонструє процеси взаємодії, де ключовими акторами виступають звичайний користувач, експерт та зовнішня система YouTube Data API v3, яка забезпечує збір необхідних для аналізу даних.

Звичайний користувач взаємодіє із системою шляхом подання запиту на аналіз, введення URL або ідентифікатора відео чи каналу, перегляду результатів аналізу та експорту сформованого звіту. Для виконання аналізу система отримує дані про відео, канали та коментарі через YouTube Data API v3.

Експерт бере участь у процесі перевірки результатів класифікації та може додавати експертний висновок до сформованого звіту. Зовнішня система YouTube Data API v3 забезпечує надання даних, необхідних для формування поведінкових характеристик користувачів, виявлення аномалій та подальшої

класифікації облікових записів. Таким чином, діаграма прецедентів відображає основні функціональні можливості системи та взаємодію її акторів під час виконання процесу виявлення ботоподібної активності.

Для детальнішого відображення логіки функціонування системи була побудована діаграма діяльності, яка описує послідовність виконання операцій під час аналізу облікового запису на платформі YouTube. На діаграмі відображено процес обробки запиту користувача, отримання даних через YouTube Data API v3, формування поведінкових характеристик, виявлення аномалій, класифікацію облікових записів і формування підсумкового звіту. Також враховано альтернативні сценарії, пов’язані з перевіркою коректності URL, використанням кешованих результатів та обробкою помилок отримання даних. Діаграма діяльності процесу виявлення ботоподібної активності на платформі YouTube наведена у додатку Б.

Як показано на діаграмі діяльності, процес аналізу починається з введення користувачем URL відео або каналу та перевірки його коректності. Після успішної перевірки система виділяє ідентифікатор ресурсу та перевіряє наявність готового результату в кеші. Якщо звіт уже існує, він одразу надається користувачу. В іншому випадку виконується запит до YouTube Data API v3 і збір даних для подальшого аналізу.

Після отримання даних здійснюється їх обробка, формування набору ознак, виявлення аномальної поведінки та класифікація облікових записів за допомогою алгоритмів машинного навчання. На завершальному етапі система формує і зберігає звіт, відображає результати користувачу та, за потреби, забезпечує його експорт. Така послідовність дій дає змогу оптимізувати використання ресурсів системи та скоротити час повторного аналізу завдяки механізму кешування[15].

У підрозділі змодельовано функціональні вимоги до інтелектуальної системи виявлення ботоподібної активності на платформі YouTube за допомогою UML-діаграм. Діаграма прецедентів визначає основних акторів системи та сценарії їх взаємодії з програмним забезпеченням. Діаграма

діяльності відображає послідовність аналізу облікових записів і формування результатів перевірки. Отримані результати слугують основою для подальшого проєктування структури та архітектури системи.

2.2 Проєктування структури та архітектури системи

Проєктування системи виявлення ботоподібної активності в YouTube є важливим етапом розробки, який передбачає визначення основних функціональних компонентів системи та їх взаємодії між собою. На даному етапі формується структура системи, визначаються потоки даних та послідовність виконання процесів, необхідних для аналізу користувацької активності, виявлення аномальної поведінки та класифікації облікових записів.

Для опису функціональної структури системи було використано методологію функціонального моделювання IDEF0. Даний підхід дозволяє представити процес роботи системи у вигляді сукупності взаємопов'язаних функціональних блоків та визначити потоки даних між ними. У межах функціонального моделювання було виконано декомпозицію основного процесу виявлення ботоподібної активності на окремі підпроцеси, що забезпечують послідовне виконання аналізу від моменту надходження запиту користувача до формування підсумкового звіту.

Декомпозиція функціональної моделі інтелектуальної системи виявлення ботоподібної активності на платформі YouTube наведена на рисунку 2.1.

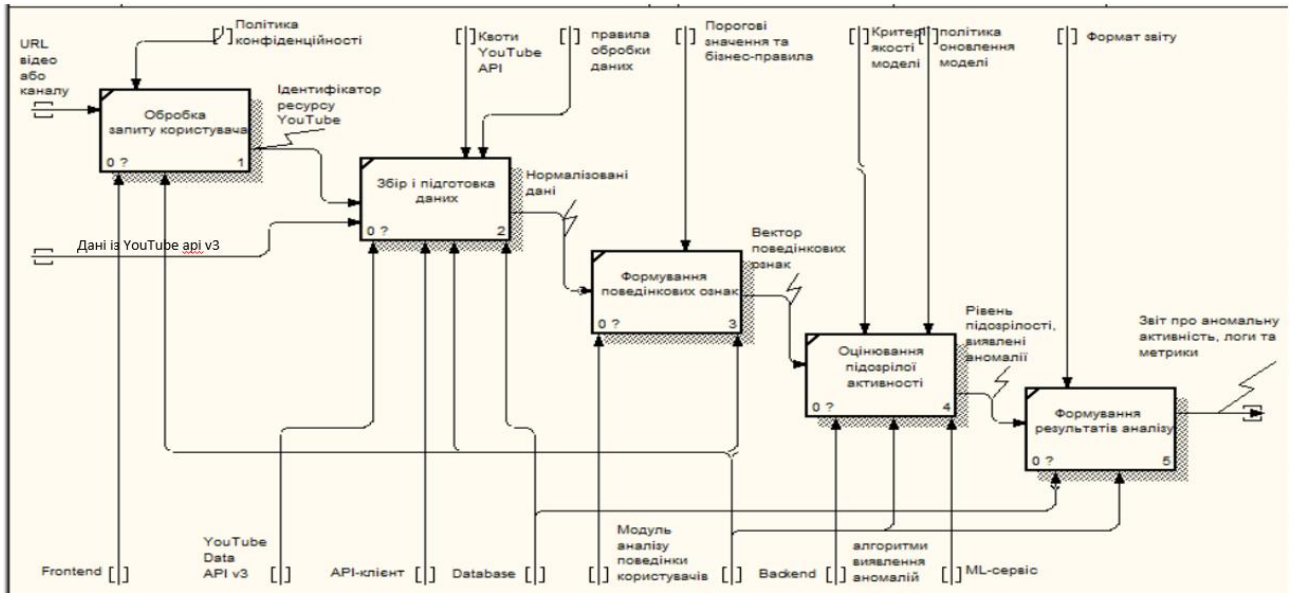


Рисунок 2.1 – Контекстна діаграма IDEF0 інтелектуальної системи виявлення ботоподібної активності на платформі YouTube.

Як показано на рисунку 2.1, процес виявлення ботоподібної активності складається з п’яти основних функціональних блоків. На першому етапі виконується обробка запиту користувача, під час якої здійснюється перевірка та виділення ідентифікатора ресурсу YouTube. Далі система виконує збір і підготовку даних за допомогою YouTube Data API v3, після чого формуються поведінкові ознаки користувачів на основі отриманої інформації[18].

Сформовані ознаки використовуються для виявлення аномальної та ботоподібної активності користувачів. Результати аналізу передаються до блоку формування підсумкового звіту, збереження журналів роботи системи та службових метрик. Така декомпозиція дозволяє чітко структурувати процес аналізу YouTube-коментарів і визначити взаємозв’язки між основними компонентами системи.

2.3 Проєктування бази даних та програмних компонентів

Після визначення функціональної структури інтелектуальної системи виявлення ботоподібної активності на платформі YouTube необхідно деталізувати її внутрішню організацію на рівні даних та програмних компонентів. Для реалізації функцій збору, обробки та аналізу інформації система повинна забезпечувати зберігання результатів перевірок, підтримку

взаємодії між окремими модулями та можливість подальшого розширення функціональних можливостей.

Важливим етапом проєктування є розробка структури бази даних та визначення основних програмних компонентів системи. База даних забезпечує централізоване зберігання інформації, необхідної для роботи системи, тоді як програмні компоненти реалізують механізми отримання даних з платформи YouTube, формування поведінкових характеристик користувачів, виявлення аномальної активності та класифікації облікових записів. На цьому етапі визначаються основні сутності предметної області, їх взаємозв'язки та структура програмних модулів, необхідних для реалізації поставлених завдань.

Для зберігання результатів аналізу та службової інформації системи було спроєктовано базу даних, структура якої орієнтована на опрацювання відео, запусків аналізу, авторів коментарів, самих коментарів, часових сегментів активності, кластерів користувачів та підсумкових звітів. Такий підхід дозволяє зберігати не лише кінцевий результат класифікації, а й проміжні дані, необхідні для пояснення результатів роботи системи.

ER-діаграма бази даних інтелектуальної системи виявлення ботоподібної активності на платформі YouTube наведена у додатку В.

Основними сутностями бази даних є `videos`, `analysis_runs`, `authors`, `comments`, `author_edges`, `clusters`, `temporal_segments` та `reports`. Таблиця `videos` призначена для зберігання інформації про відео YouTube, що аналізуються системою. Таблиця `analysis_runs` фіксує окремі запуски аналізу та містить інформацію про статус виконання, загальний рівень ризику, рівень впевненості моделі, час початку та завершення перевірки.

Таблиця `authors` використовується для зберігання даних про авторів коментарів у межах конкретного запуску аналізу. Вона містить ідентифікатор каналу автора, ім’я, кількість коментарів, рівень ризику, оцінку підозрілості, впевненість моделі та пояснення результатів аналізу. Таблиця `comments` зберігає коментарі користувачів, їх нормалізований текст, дату публікації,

кількість вподобань, ознаку відповіді, оцінку ризику та набір ознак, сформованих під час аналізу.

Для опису взаємозв'язків між авторами використовується таблиця `author_edges`, яка дозволяє зберігати інформацію про зв'язки між підозрілими авторами та вагу цих зв'язків. Таблиця `clusters` призначена для зберігання результатів групування користувачів або коментарів за схожими поведінковими характеристиками. Таблиця `temporal_segments` використовується для аналізу активності в межах окремих часових інтервалів, що дає змогу виявляти сплески підозрілої активності. Підсумкові результати аналізу зберігаються у таблиці `reports`, яка містить загальний рівень ризику, рівень впевненості, текстовий підсумок та структурований звіт у форматі JSON.

Для опису програмної структури інтелектуальної системи було побудовано UML-діаграму класів. Вона дозволяє визначити основні програмні компоненти системи, їхні функції та зв'язки між об'єктами, які використовуються під час виконання процесу аналізу. Діаграма класів інтелектуальної системи виявлення ботоподібної активності на платформі YouTube наведена у додатку Г.

Система складається з кількох основних компонентів: обробки запиту користувача, перевірки URL, збору даних з YouTube, аналізу поведінки користувачів, виявлення аномалій, оцінювання ризику, формування звіту та збереження результатів у базі даних. Такий поділ дозволяє показати логіку роботи системи без детального опису кожного програмного класу.

Клас `BehaviorAnalyzer` використовується для формування поведінкових характеристик авторів коментарів. На основі отриманих ознак клас `AnomalyDetector` виконує виявлення аномальної активності, а `BotRiskEvaluator` визначає рівень ризику та присвоює відповідний рівень підозрілості. Клас `ReportGenerator` формує підсумковий звіт за результатами аналізу, а `DatabaseService` забезпечує збереження відео, запусків аналізу, авторів, коментарів та результатів перевірки в базі даних.

Окрему групу класів становлять сутності предметної області: `Video`, `AnalysisRun`, `Author` та `Comment`. Вони відображають основні об’єкти, з якими працює система під час аналізу. Один об’єкт `Video` може мати багато запусків аналізу, кожен запуск аналізу пов’язаний із багатьма авторами коментарів, а кожен автор може мати декілька коментарів. Така структура класів відповідає логіці роботи системи та забезпечує зв’язок між процесами збору даних, аналізу, класифікації та формування результатів.

Для відображення взаємодії між компонентами системи під час виконання аналізу було побудовано UML-діаграму послідовності. Вона демонструє порядок обміну повідомленнями між користувачем, клієнтською частиною, серверною частиною, зовнішнім сервісом YouTube Data API v3, модулем аналізу даних та сховищем даних. Діаграма послідовності інтелектуальної системи виявлення ботоподібної активності на платформі YouTube наведена у додатку Д.

Як показано на діаграмі, процес аналізу починається з введення користувачем URL відео та надсилання відповідного запиту до серверної частини системи. Після отримання запиту сервер виконує перевірку наявності раніше сформованого результату в базі даних. Якщо результат уже існує, він одразу повертається користувачу без повторного виконання аналізу. У разі відсутності результату сервер звертається до YouTube Data API v3 для отримання інформації про відео та коментарі.

Отримані дані передаються до модуля аналізу, який виконує формування поведінкових ознак, виявлення аномальної активності та оцінювання рівня ризику облікових записів. Після завершення аналізу результати зберігаються у сховищі даних та повертаються користувачу через клієнтський інтерфейс. Така послідовність взаємодії забезпечує повторне використання раніше сформованих результатів, зменшує кількість звернень до зовнішнього API та підвищує ефективність роботи системи.

Висновки до розділу 2

У другому розділі було виконано проєктування інтелектуальної системи виявлення ботоподібної активності на платформі YouTube. Для моделювання функціональних вимог використано UML-діаграму прецедентів та діаграму діяльності, що дозволило визначити основних акторів системи, сценарії їх взаємодії та послідовність виконання процесу аналізу облікових записів.

Для опису функціональної структури системи було застосовано методологію IDEF0, за допомогою якої процес виявлення ботоподібної активності в YouTube-коментарях поділено на окремі функціональні блоки: обробку запиту користувача, збір і підготовку даних, формування поведінкових ознак, виявлення аномальної активності та формування результатів аналізу.

Також було спроєктовано структуру бази даних системи, яка забезпечує зберігання інформації про відео, запуски аналізу, авторів коментарів, коментарі, часові сегменти, кластери, зв’язки між авторами та підсумкові звіти. Для опису програмної структури системи побудовано UML-діаграму класів, а для відображення взаємодії між компонентами під час виконання аналізу – діаграму послідовності.

Отримані результати проєктування є основою для подальшої програмної реалізації інтелектуальної системи, розробки її основних модулів та проведення тестування працездатності.

РОЗДІЛ 3. РОЗРОБКА ТА РЕАЛІЗАЦІЯ ІНТЕЛЕКТУАЛЬНОЇ СИСТЕМИ

3.1 Формування поведінкових ознак користувачів

Ефективність виявлення ботоподібної активності значною мірою залежить від якості сформованих ознак, які описують особливості поведінки користувачів на платформі YouTube. Первинні дані, отримані через YouTube Data API v3, містять інформацію про коментарі, авторів коментарів, часові характеристики публікацій та статистичні показники активності. Однак у такому вигляді вони не можуть бути безпосередньо використані для аналізу, тому потребують попередньої обробки та перетворення у набір формалізованих характеристик.

У розроблюваній інтелектуальній системі формування ознак виконується шляхом аналізу текстового вмісту коментарів, часових параметрів активності користувачів та поведінкових характеристик авторів. Отримані ознаки використовуються на наступних етапах роботи системи для виявлення аномальної активності, аналізу схожості коментарів та формування інтегральної оцінки ризику ботоподібної поведінки.

Одним із основних джерел інформації для виявлення ботоподібної активності є текстовий вміст коментарів. Перед виконанням аналізу всі отримані повідомлення проходять етап попередньої обробки, який включає приведення тексту до єдиного регістру, видалення надлишкових пробілів та нормалізацію структури повідомлення. Такий підхід дозволяє зменшити вплив відмінностей у форматуванні тексту на результати подальшого аналізу.

Після нормалізації для кожного коментаря формується набір текстових ознак: довжина повідомлення, кількість слів і токенів, наявність посилань, емодзі, цифр, пунктуації та повторюваних символів. Ці характеристики дозволяють виявляти коментарі зі спамоподібною структурою та використовуються для подальшого оцінювання ризику підозрілої активності.

Важливою характеристикою є також оцінка лексичної різноманітності повідомлення. Для цього аналізується співвідношення між кількістю унікальних

слів та загальною кількістю слів у коментарі. Низька різноманітність тексту може свідчити про використання шаблонних повідомлень або автоматичну генерацію контенту.

Для кількісного оцінювання лексичної різноманітності використовується коефіцієнт унікальності токенів:

$$Unique\ Token\ Ratio = \frac{|unique(tokens)|}{|tokens|} \quad (3.1)$$

де (|unique(tokens)|) - кількість унікальних слів у коментарі, а (tokens) - загальна кількість слів. Чим менше значення показника, тим більш шаблонним є текст повідомлення, що може свідчити про використання автоматизованих засобів генерації контенту.

Крім цього, система виконує словниковий пошук типових спамових конструкцій, закликів до дії та підозрілих ключових слів. Аналізується наявність фраз, пов’язаних із рекламою, просуванням сторонніх ресурсів або переходом за посиланнями. Такі ознаки використовуються як додаткові сигнали під час оцінювання ризику підозрілої активності.Окрім текстового аналізу, важливим джерелом інформації є часові характеристики активності користувачів. Аналіз часових параметрів дозволяє виявляти нетипові шаблони поведінки, які часто характерні для ботів або координованих груп акаунтів.

Для кожного коментаря визначається час його публікації та інтервали між сусідніми повідомленнями. На основі цих даних обчислюється інтенсивність активності користувачів, щільність появи коментарів у часових інтервалах та частота повторних публікацій.

Особлива увага приділяється виявленню короткочасних сплесків активності. Подібні ситуації можуть виникати як внаслідок природної реакції аудиторії на популярний контент, так і в результаті координованої діяльності бот-мереж. Для розмежування цих випадків система аналізує кількість активних авторів, повторюваність їх участі у часових сегментах та ступінь концентрації повідомлень у визначені проміжки часу.

Для оцінювання інтенсивності активності у часовому сегменті використовується коефіцієнт сплеску активності:

$$burst_ratio = \frac{n_b}{global_burst_baseline} \quad (3.3)$$

де n_b - кількість коментарів у поточному часовому сегменті, а $global_burst_baseline$ - середня кількість коментарів у всіх сегментах. Зростання даного показника може свідчити про наявність координованої активності користувачів.

Додатково оцінюється рівномірність розподілу активності у часі та виявляються аномальні відхилення від типових шаблонів поведінки. Отримані характеристики використовуються для подальшого виявлення підозрілої активності та формування оцінки ризику.

Важливим етапом аналізу є формування поведінкового профілю автора коментарів. На відміну від окремих повідомлень, поведінковий профіль дозволяє оцінити сукупну активність користувача та виявити довгострокові закономірності його поведінки.

Для кожного автора система визначає загальну кількість коментарів, частоту їх публікації, рівень повторюваності повідомлень та інші статистичні характеристики. Окремо аналізується використання однакових або дуже схожих текстів у різних повідомленнях, що може свідчити про застосування шаблонів або автоматизованих засобів генерації контенту.

До поведінкового профілю також включаються показники використання посилань, коротких повідомлень та інших характеристик, отриманих під час текстового аналізу. Додатково враховується вік облікового запису та інтенсивність його активності. Молоді акаунти з високою частотою публікації однотипних коментарів можуть розглядатися як потенційно підозрілі, особливо за наявності інших ознак автоматизованої поведінки.

Сформований набір текстових, часових та поведінкових ознак утворює основу для подальшого аналізу. На наступному етапі система використовує отримані характеристики для виявлення аномальної активності, аналізу

взаємозв'язків між користувачами та формування інтегральної оцінки ризику ботоподібної поведінки.

3.2 Виявлення аномальної поведінки та оцінювання ризику

Після формування текстових, часових та поведінкових ознак система переходить до етапу виявлення аномальної активності користувачів. Основною метою даного етапу є виявлення нетипових шаблонів поведінки, які можуть свідчити про використання автоматизованих облікових записів або координовану діяльність груп користувачів.

На відміну від традиційних підходів, що базуються на окремих правилах або порогових значеннях, у розроблюваній системі використовується комплексний аналіз декількох груп ознак. Це дозволяє враховувати текстову активність користувачів, часові характеристики публікації коментарів, схожість повідомлень та взаємозв'язки між авторами.

Першим етапом виявлення аномальної активності є аналіз текстового вмісту коментарів. Система виконує пошук коротких, майже порожніх або шаблонних повідомлень, а також визначає наявність рекламних конструкцій та підозрілих ключових слів. Додатково здійснюється виявлення однакових та схожих повідомлень, що можуть свідчити про використання автоматизованих сценаріїв поширення коментарів. Для цього застосовується метод TF-IDF та косинусна міра схожості, що дозволяє об'єднувати подібні повідомлення у групи та визначати ступінь їх повторюваності[4].

Оцінювання схожості між коментарями виконується за допомогою косинусної міри схожості:

$$sim(i,j) = \frac{v_i \cdot v_j}{||v_i|| \cdot ||v_j||} \tag{3.2}$$

де v_i та v_j — TF-IDF вектори двох коментарів. Значення показника знаходиться в діапазоні від 0 до 1, де більші значення відповідають більшій текстовій схожості повідомлень.

Другим етапом є аналіз часових характеристик активності користувачів. Система досліджує інтервали між коментарями, щільність їх появи у часових сегментах та інтенсивність активності окремих авторів. Особлива увага приділяється виявленню короткочасних сплесків активності, які можуть бути ознакою координованої роботи бот-мереж. Для кожного часового сегмента оцінюється рівень синхронності дій користувачів, концентрація повідомлень та повторна участь авторів у підозрілих часових вікнах.

Наступним етапом є аналіз поведінкових профілів користувачів. Для кожного автора формується набір характеристик, що включає кількість коментарів, частоту їх публікації, рівень дублювання повідомлень, використання посилань, рекламних конструкцій та інших поведінкових ознак. Отримані дані використовуються для порівняння поведінки різних користувачів та виявлення нетипових профілів активності.

Для пошуку авторів із подібними поведінковими характеристиками застосовується алгоритм кластеризації DBSCAN. На вхід алгоритму подається набір стандартизованих авторських ознак, що описують частоту коментування, повторюваність повідомлень, використання посилань, коротких текстів та інші показники активності. Перевагою DBSCAN є те, що він не потребує попереднього задання кількості кластерів і дозволяє виділяти як групи користувачів зі схожою поведінкою, так і окремі аномальні об’єкти[8].

Окрім поведінкової кластеризації, у системі використовується графовий аналіз взаємозв’язків між авторами коментарів. Для цього формується граф, у якому вершинами є автори, а ребрами — виявлені ознаки можливої координації між ними. Під час побудови графа враховуються схожість текстів коментарів, синхронність активності, участь авторів у спільних підозрілих часових сегментах, а також використання однакових або подібних URL-доменів. Після відкидання слабких зв’язків аналіз структури графа дозволяє виділити групи авторів, між якими спостерігаються ознаки координованої діяльності.

Такий підхід дозволяє аналізувати не лише окремі акаунти, а й взаємодію між ними. Це важливо, оскільки ботоподібна активність часто проявляється не в

ізольованій поведінці одного користувача, а у спільних діях групи акаунтів: одночасному розміщенні коментарів, використанні схожих повідомлень або поширенні однакових посилань. Тому графовий аналіз доповнює результати кластеризації та підвищує обґрунтованість висновку про наявність координованої активності.

Для виявлення нетипових поведінкових профілів застосовується алгоритм Isolation Forest. Даний метод належить до алгоритмів пошуку аномалій без використання попередньо розмічених даних та дозволяє визначати об'єкти, поведінка яких суттєво відрізняється від поведінки більшості користувачів. У процесі аналізу враховуються кількість коментарів, частка дубльованих повідомлень, використання посилань, інтенсивність коментування та інші поведінкові характеристики. Результатом роботи алгоритму є оцінка аномальності користувача, яка використовується на наступному етапі аналізу.

Отримані результати текстового аналізу, часового аналізу, поведінкової кластеризації, графового аналізу та виявлення аномалій об'єднуються для формування комплексної оцінки активності користувача. Такий підхід дозволяє зменшити кількість хибних спрацьовувань та підвищити достовірність виявлення ботоподібної поведінки. На основі сформованих оцінок виконується подальше визначення рівня ризику, що розглядається у наступному підрозділі.

3.3 Оцінювання ризику ботоподібної поведінки користувачів

Після завершення етапів формування ознак та виявлення аномальної активності система переходить до оцінювання ризику ботоподібної поведінки користувачів. Основною метою цього етапу є інтеграція результатів різних методів аналізу та формування єдиної узагальненої оцінки, яка характеризує ймовірність використання автоматизованого облікового запису або участі користувача в координованій активності.

На відміну від традиційних підходів, що приймають рішення на основі одного алгоритму або окремого показника, у розроблюваній системі використовується багаторівневий механізм інтеграції ризиків[29]. Такий підхід

дозволяє зменшити кількість хибних спрацьовувань та забезпечити більш обґрунтоване прийняття рішень.

Для кожного автора система формує декілька незалежних груп сигналів ризику. До них належать часові сигнали, що характеризують особливості активності користувача у часі; координаційні сигнали, які відображають зв'язки між авторами; поведінкові сигнали, сформовані на основі аналізу текстових та статистичних характеристик; сигнали повторюваності активності між різними запусками аналізу; а також сигнали текстової схожості, що відображають рівень дублювання або шаблонності коментарів.

На першому етапі для кожної групи ознак формується окрема оцінка ризику. Часова оцінка враховує сплески активності, повторну участь автора у підозрілих часових сегментах та інтенсивність коментування. Координаційна оцінка базується на аналізі зв'язків між користувачами та рівні синхронності їхньої поведінки. Поведінкова оцінка враховує дублювання повідомлень, використання посилань, підозрілих ключових слів та інші характеристики текстової активності. Оцінка повторюваності відображає сталість підозрілих шаблонів поведінки протягом декількох аналізів. Додатково враховується рівень текстової схожості між повідомленнями різних користувачів.

Після формування окремих оцінок система визначає ступінь їх узгодженості. Важливою особливістю реалізованого підходу є те, що наявність лише одного сильного сигналу не вважається достатньою підставою для формування високого ризику. Для підвищення достовірності результатів система аналізує кількість незалежних груп ознак, які підтверджують підозрілу активність. Чим більше незалежних джерел підтверджують наявність аномалій, тим вищою є підсумкова оцінка ризику.

Додатково застосовується механізм обмеження ризику для випадків, коли підозра ґрунтується лише на одному типі ознак. Наприклад, якщо користувач має лише нетипову часову активність або лише високий рівень текстової схожості повідомлень, система знижує максимальну оцінку ризику до отримання

додаткових підтверджень від інших модулів аналізу. Такий підхід дозволяє уникати помилкового визначення активних користувачів як ботів.

При формуванні підсумкової оцінки також враховуються додаткові фактори. До них належать вік облікового запису, результати пошуку аномалій за допомогою алгоритму Isolation Forest, результати поведінкової кластеризації DBSCAN, а також показники різноманітності коментарів. Дані характеристики використовуються як допоміжні сигнали та не можуть самостійно формувати високий рівень ризику без підтвердження з боку інших груп ознак[5].

Результатом роботи модуля є підсумкова оцінка ризику, рівень ризику та внутрішня оцінка впевненості системи у сформованому результаті. Залежно від значення підсумкової оцінки користувач може бути віднесений до категорії низького, середнього або високого ризику. Додатково система формує пояснення щодо основних факторів, які вплинули на прийняте рішення, що підвищує прозорість роботи системи та спрощує інтерпретацію результатів аналізу.

Отримані оцінки використовуються для формування фінального звіту, який відображається користувачу та містить інформацію про виявлені аномалії, рівень ризику та фактори, що вплинули на підсумковий висновок системи.

3.4 Програмна реалізація та інтерфейс користувача

Розроблена інтелектуальна система реалізована у вигляді вебзастосунку, що забезпечує взаємодію користувача з модулями збору даних, аналізу поведінки та виявлення підозрілої активності. Основним призначенням інтерфейсу є надання зручного доступу до функцій аналізу відео або каналів платформи YouTube та відображення результатів роботи системи у зрозумілому для користувача вигляді.

Інтерфейс побудовано з використанням сучасних вебтехнологій та орієнтовано на покрокову взаємодію користувача із системою. Після запуску користувач вводить посилання на відео або канал, після чого система автоматично виконує збір даних, формування ознак, виявлення аномалій та оцінювання ризику підозрілої активності.

Головна сторінка є початковою точкою взаємодії користувача із системою. Вона містить короткий опис призначення програмного продукту, інформацію про стан сервісу та форму запуску аналізу.

На сторінці реалізовано вибір режиму роботи системи. Користувач може виконати аналіз окремого відео або каналу. Для запуску аналізу необхідно ввести URL-адресу ресурсу YouTube або його ідентифікатор у відповідне поле вводу. Після перевірки коректності введених даних користувач може запустити процедуру аналізу натисканням відповідної кнопки.

Окремим елементом інтерфейсу є блок контролю стану сервісу, який дозволяє відображати готовність системи до виконання аналізу та повідомляти користувача про доступність сервісу.

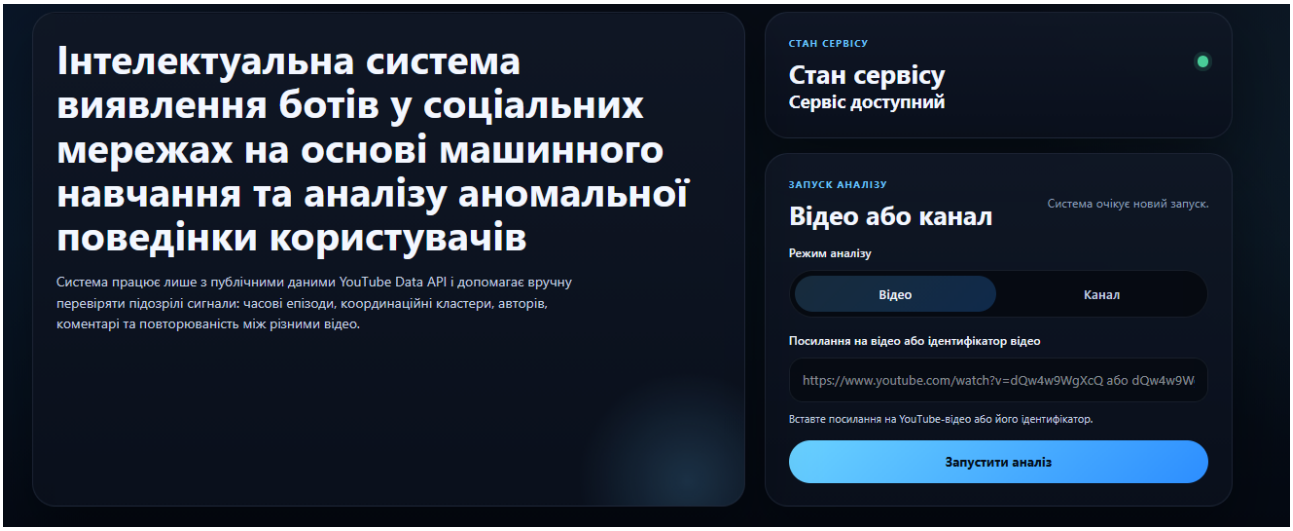


Рисунок 3.2 – Головна сторінка системи

Після завершення обробки даних система формує підсумковий результат аналізу та відображає його користувачу у вигляді інтегральної оцінки ризику.

У верхній частині сторінки відображається значення підсумкового ризику підозрілої активності у відсотках, а також рівень впевненості сформованого висновку. Додатково система надає коротке текстове пояснення щодо виявлених ознак, які вплинули на результат аналізу.

Для спрощення інтерпретації результатів користувачу надається узагальнена статистика, що містить кількість проаналізованих коментарів, підозрілих авторів, часових епізодів та координаційних кластерів. Такий підхід

дозволяє швидко оцінити загальний рівень підозрілої активності та перейти до детального аналізу окремих елементів.

Оскільки система виконує оцінювання ризику, а не автоматичну класифікацію користувачів, в інтерфейсі додатково відображаються застереження щодо інтерпретації результатів. Це дозволяє уникнути помилкового трактування отриманих висновків як остаточного підтвердження наявності ботів.

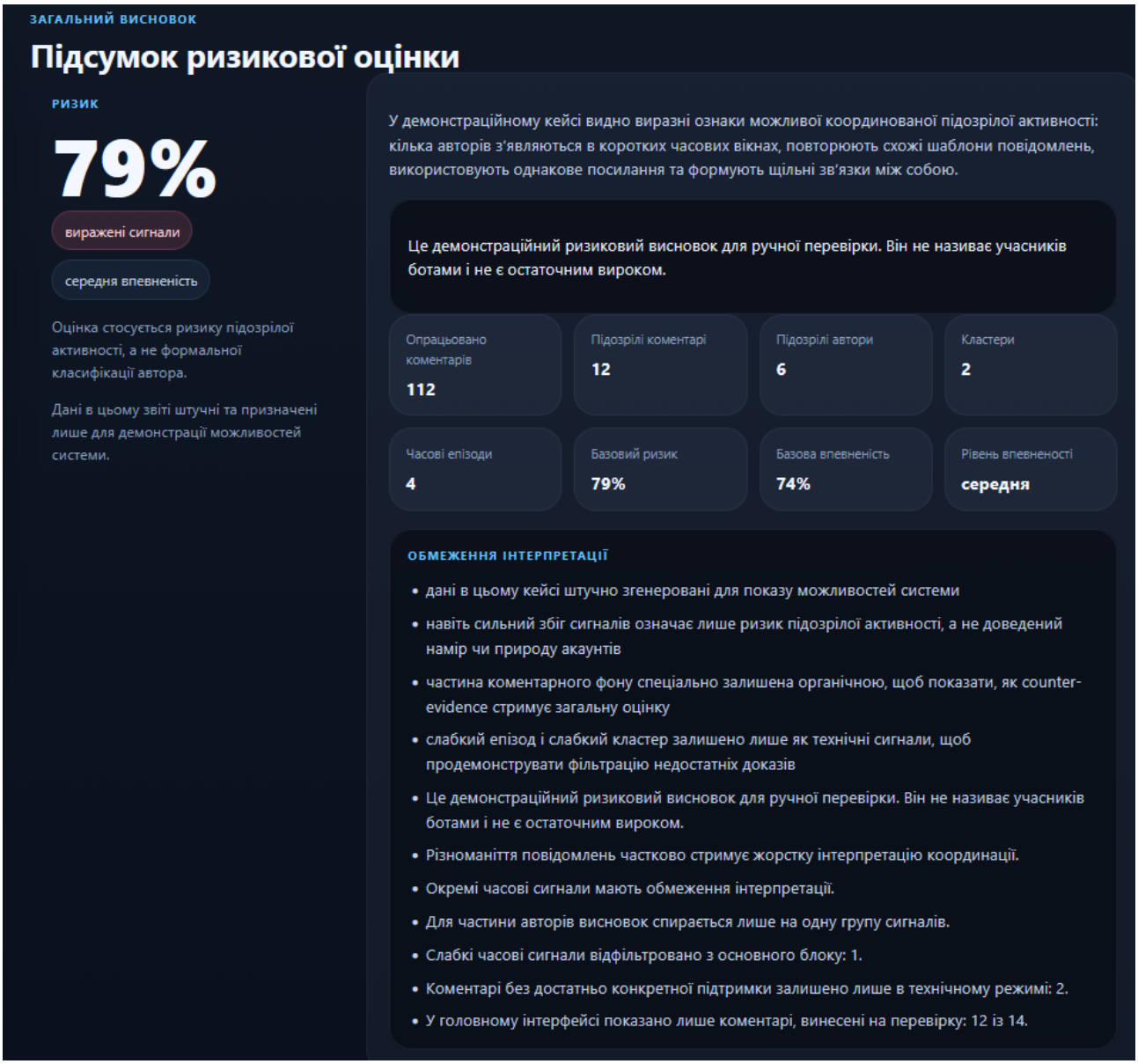


Рисунок 3.3 – Загальний результат аналізу

Одним із важливих компонентів інтерфейсу є модуль візуалізації часових аномалій. Його основним завданням є відображення підозрілих часових сегментів, у межах яких спостерігається нетипова активність користувачів.

Для кожного виявленого епізоду система відображає часовий інтервал, кількість коментарів, кількість активних авторів та рівень ризику. Візуальне представлення у вигляді шкали дозволяє швидко оцінити відносний рівень підозрілості кожного сегмента.

Використання часової шкали спрощує аналіз динаміки активності користувачів та дозволяє виявляти короткочасні сплески коментування, характерні для координованих інформаційних кампаній або автоматизованих сценаріїв поширення повідомлень.

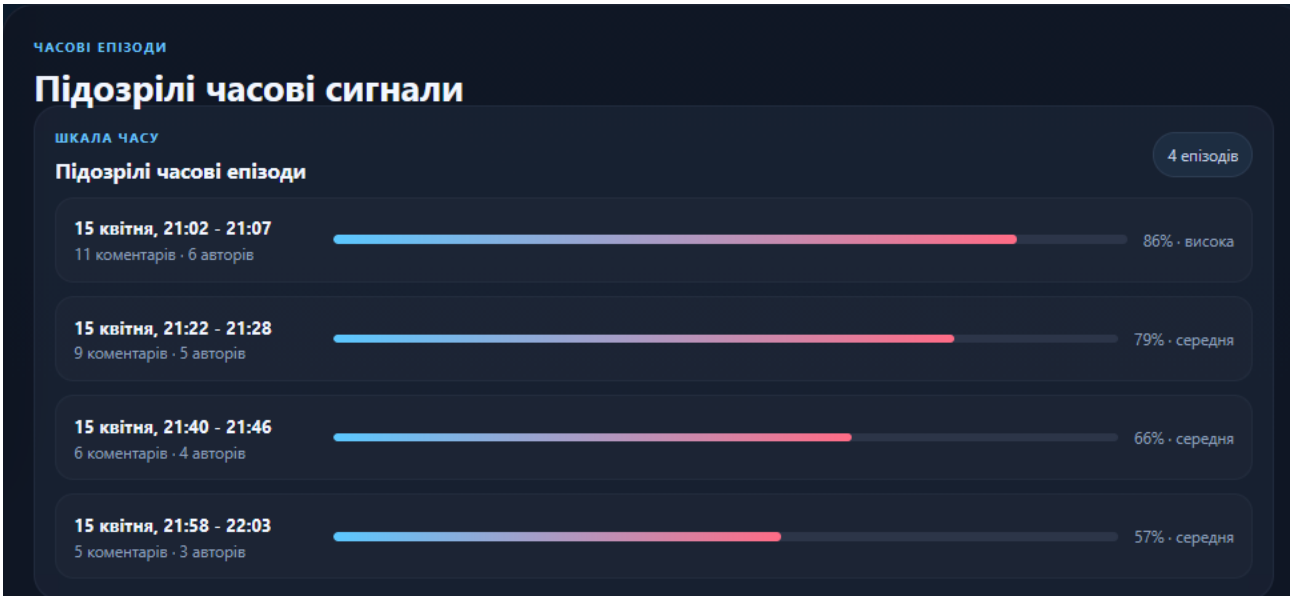


Рисунок 3.4 – Виявлені часові епізоди

Для кожного підозрілого епізоду користувач може перейти до детального перегляду. У такому режимі відображаються кількість учасників події, рівень ризику, рівень впевненості та список коментарів, які були враховані під час формування висновку.

Додатково система надає текстове пояснення причин формування ризикової оцінки та перелік ознак, що підтверджують наявність підозрілої активності. Це дозволяє виконувати ручну перевірку результатів та оцінювати обґрунтованість автоматично сформованих висновків.

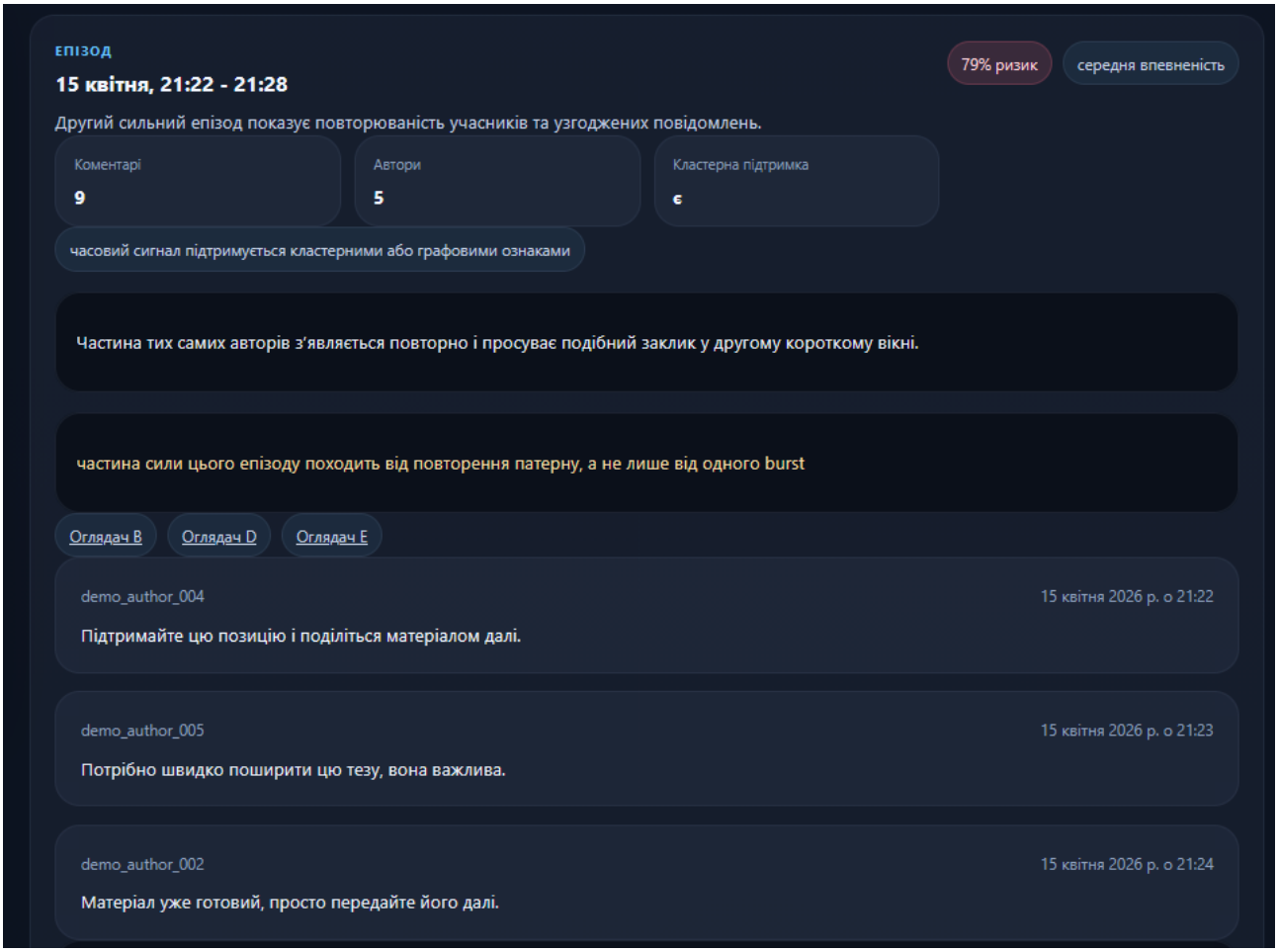


Рисунок 3.5 – Детальна інформація про часовий епізод

Для аналізу груп користувачів зі схожою поведінкою в системі реалізовано окремий модуль відображення координаційних кластерів.

У межах цього модуля відображаються групи авторів, між якими було виявлено ознаки узгодженої діяльності[12]. Для кожного кластера система надає інформацію про кількість учасників, рівень ризику, рівень впевненості, кількість синхронних часових вікон та основні фактори, що стали підставою для формування висновку.

Окремо відображаються текстові пояснення, які допомагають користувачу зрозуміти характер виявлених зв'язків між учасниками групи. До таких пояснень можуть належати повідомлення про використання схожих текстових шаблонів, спільних посилань, повторюваних часових шаблонів активності або графових зв'язків між авторами.

Важливою особливістю інтерфейсу є відображення застережень щодо необхідності ручної перевірки результатів кластерного аналізу. Навіть за

наявності сильних сигналів координації система не формує автоматичний висновок про природу облікових записів, а лише надає інформацію для подальшого аналізу.

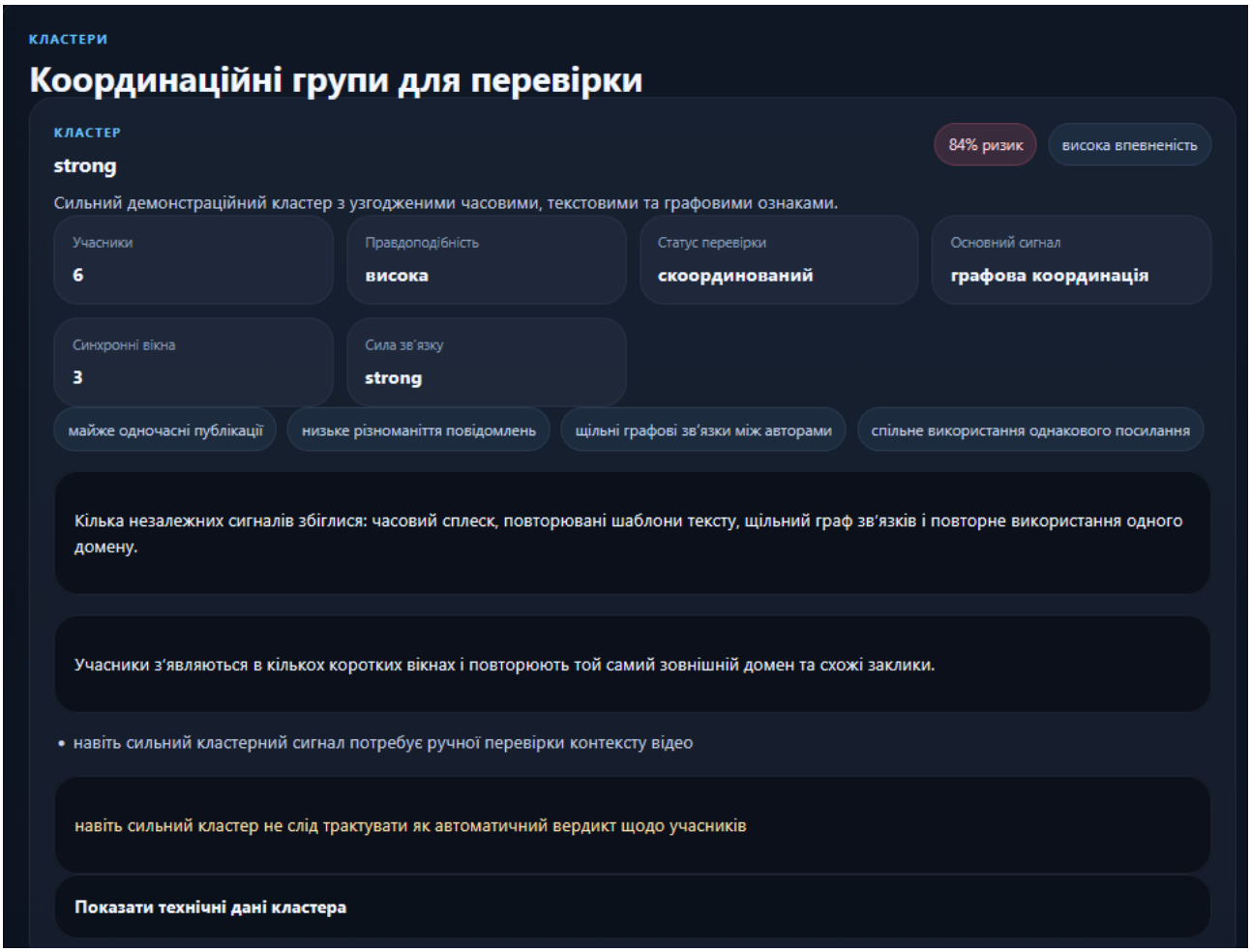


Рисунок 3.6 – Координаційний кластер користувачів

Для кожного автора, щодо якого було виявлено підозрілі сигнали, система формує окрему інформаційну картку.

У картці відображаються підсумкова оцінка ризику, рівень впевненості, домінуючий сигнал ризику, кількість незалежних підтверджуючих ознак та рівень узгодженості між різними групами сигналів. Також користувачу надається перелік факторів, які вплинули на формування остаточного висновку.

Для підвищення прозорості роботи системи відображаються числові значення окремих показників, зокрема часових, координаційних та поведінкових сигналів. Це дозволяє користувачу зрозуміти причини підвищення або зниження ризикової оцінки.

Додатково система формує текстові пояснення щодо сильних і слабких сторін доказової бази, а також відображає інформацію про фактори, які могли вплинути на підсумковий результат.

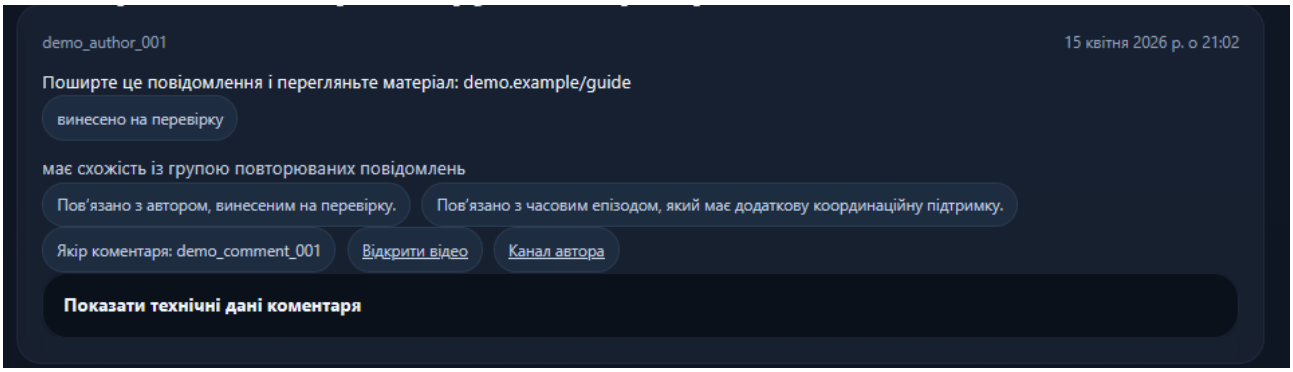


Рисунок 3.7 – Результати аналізу автора

Отже, розроблений інтерфейс забезпечує зручну взаємодію користувача із системою та дозволяє послідовно переходити від загальної оцінки ризику до детального аналізу окремих авторів, часових епізодів і координаційних груп. Реалізовані механізми візуалізації результатів підвищують зрозумілість роботи системи та спрощують процес інтерпретації отриманих висновків.

Висновки до розділу 3

У третьому розділі було розглянуто особливості функціонування та програмної реалізації інтелектуальної системи виявлення ботоподібної активності на платформі YouTube. Описано процес формування поведінкових ознак користувачів на основі аналізу текстового вмісту коментарів, часових характеристик активності та поведінкових профілів авторів. Визначено набір ознак, що використовується для подальшого аналізу та виявлення підозр

ілої активності.

Розглянуто механізм виявлення аномальної поведінки користувачів із застосуванням методів аналізу текстової схожості, часових закономірностей, графового аналізу взаємозв'язків та алгоритмів машинного навчання DBSCAN і Isolation Forest. Запропонований підхід дозволяє виявляти як окремі аномальні облікові записи, так і групи користувачів із ознаками координованої активності.

Описано механізм інтеграції результатів аналізу для формування комплексної оцінки ризику ботоподібної поведінки. Підсумкова оцінка ризику

враховує часові, поведінкові, координаційні та текстові сигнали, що дозволяє підвищити достовірність прийняття рішень і зменшити кількість хибних спрацьовувань.

Також представлено програмну реалізацію системи та основні елементи користувацького інтерфейсу. Розроблений вебзастосунок забезпечує введення URL-адреси відео або каналу, автоматичний запуск аналізу, відображення результатів оцінювання ризику та візуалізацію виявлених аномалій. Це забезпечує зручну взаємодію користувача із системою та спрощує інтерпретацію отриманих результатів.

Таким чином, у межах третього розділу було реалізовано та описано основні механізми роботи інтелектуальної системи виявлення ботоподібної активності на платформі YouTube, що забезпечують автоматизований аналіз поведінки користувачів, виявлення аномальної активності та формування обґрунтованої оцінки ризику ботоподібної поведінки.

ВИСНОВКИ

У кваліфікаційній роботі вирішено актуальну задачу розробки інтелектуальної системи виявлення ботоподібної активності на платформі YouTube на основі машинного навчання та аналізу аномальної поведінки користувачів.

У результаті виконання роботи було проведено аналіз предметної області та встановлено, що автоматизовані облікові записи суттєво впливають на достовірність статистичних показників, інформаційне середовище та взаємодію користувачів на платформі YouTube. Проведене дослідження існуючих методів і програмних засобів показало доцільність використання комплексного підходу, який поєднує поведінковий аналіз, методи машинного навчання та виявлення аномальної активності.

На основі проведеного аналізу було сформовано функціональні та нефункціональні вимоги до системи, визначено основні вхідні та вихідні дані, а також розроблено функціональну модель і архітектуру програмного забезпечення. Для опису структури системи використано UML-діаграми та функціональне моделювання IDEF0, що дозволило формалізувати процеси збору даних, формування ознак, виявлення аномалій та оцінювання ризику ботоподібної поведінки.

У роботі реалізовано механізм формування поведінкових ознак користувачів на основі аналізу текстового вмісту коментарів, часових характеристик активності та поведінкових профілів авторів. Сформований набір ознак використовується для подальшого аналізу та виявлення підозрілої активності користувачів.

Для пошуку аномальної поведінки використано алгоритм Isolation Forest, який дозволяє виявляти нетипові профілі активності без використання попередньо розмічених даних. Для виявлення груп користувачів зі схожою поведінкою застосовано алгоритм кластеризації DBSCAN. Додатково реалізовано аналіз текстової схожості коментарів та графовий аналіз

взаємозв'язків між авторами, що дозволяє виявляти ознаки координованої діяльності.

Розроблено механізм інтеграції результатів аналізу, який враховує часові, поведінкові, координаційні та текстові сигнали ризику. Такий підхід забезпечує формування комплексної оцінки ризику ботоподібної поведінки та сприяє зменшенню кількості хибних спрацьовувань під час аналізу користувацької активності.

Практичним результатом роботи є реалізований вебзастосунок для аналізу відео та каналів платформи YouTube. Система забезпечує автоматичний збір даних через YouTube Data API v3, формування поведінкових ознак, виявлення аномалій, оцінювання ризику підозрілої активності та відображення результатів у зручному для користувача вигляді. Реалізовані засоби візуалізації дозволяють аналізувати часові аномалії, координаційні кластери та результати перевірки окремих авторів.

Отримані результати можуть бути використані для моніторингу інформаційного простору YouTube, виявлення ботоподібної активності, аналізу координованих інформаційних кампаній та підтримки прийняття рішень під час дослідження активності користувачів у соціальних мережах. Подальшим напрямом розвитку роботи може бути розширення набору поведінкових ознак, використання додаткових алгоритмів машинного навчання та адаптація системи для аналізу інших соціальних платформ.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Аналіз технологій виявлення ботоподібної активності у соціальних мережах. *Інформаційна безпека та комп'ютерні технології*. 2023. С. 5. URL: https://kntu.kr.ua/doc/science/zahody/vikl/2023/6-tez.pdf#page=5.
2. Функціювання ботів у соцмережах як різновид тролингу. *Актуальні проблеми природничих і гуманітарних наук у дослідженнях молодих учених*. 2021. С. 634.
3. Anson. YouTube data API tutorial - search for videos, 2022. *YouTube*. URL: https://www.youtube.com/watch?v=QY8dhl1EQfI.
4. Cosine_similarity. *scikit-learn*. URL: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.pairwise.cosine_similarity.html.
5. Dbscan. *scikit-learn*. URL: https://scikit-learn.org/stable/modules/generated/sklearn.cluster.DBSCAN.html.
6. Docker Inc. Manuals. *Docker Documentation*. URL: https://docs.docker.com/manuals/.
7. Exposure to social bots amplifies perceptual biases and regulation propensity - PMC. *PMC Home*. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC10673860/.
8. FastAPI - FastAPI. *FastAPI - FastAPI*. URL: https://fastapi.tiangolo.com/.
9. Google cloud documentation. *Google Cloud Documentation*. URL: https://docs.cloud.google.com/chronicle/docs/preview/detection-engine/yara-l-2-0-functions/strings-shannon_entropy.
10. How bots affect discussions: a study of YouTube comments in March 2025. *Factcheck*. URL: https://factcheck.lt/eng/news/youtube_botnet_march2025/.
11. IsolationForest. *scikit-learn*. URL: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.IsolationForest.html.
12. Lynnette Hui Xian Ng and Kathleen M. Carley. Modeling the impact of social media bots for information dissemination. URL: https://www.cmu.edu/ideas-social-cybersecurity/events/2024_ideas_cmot_lynnette.pdf.

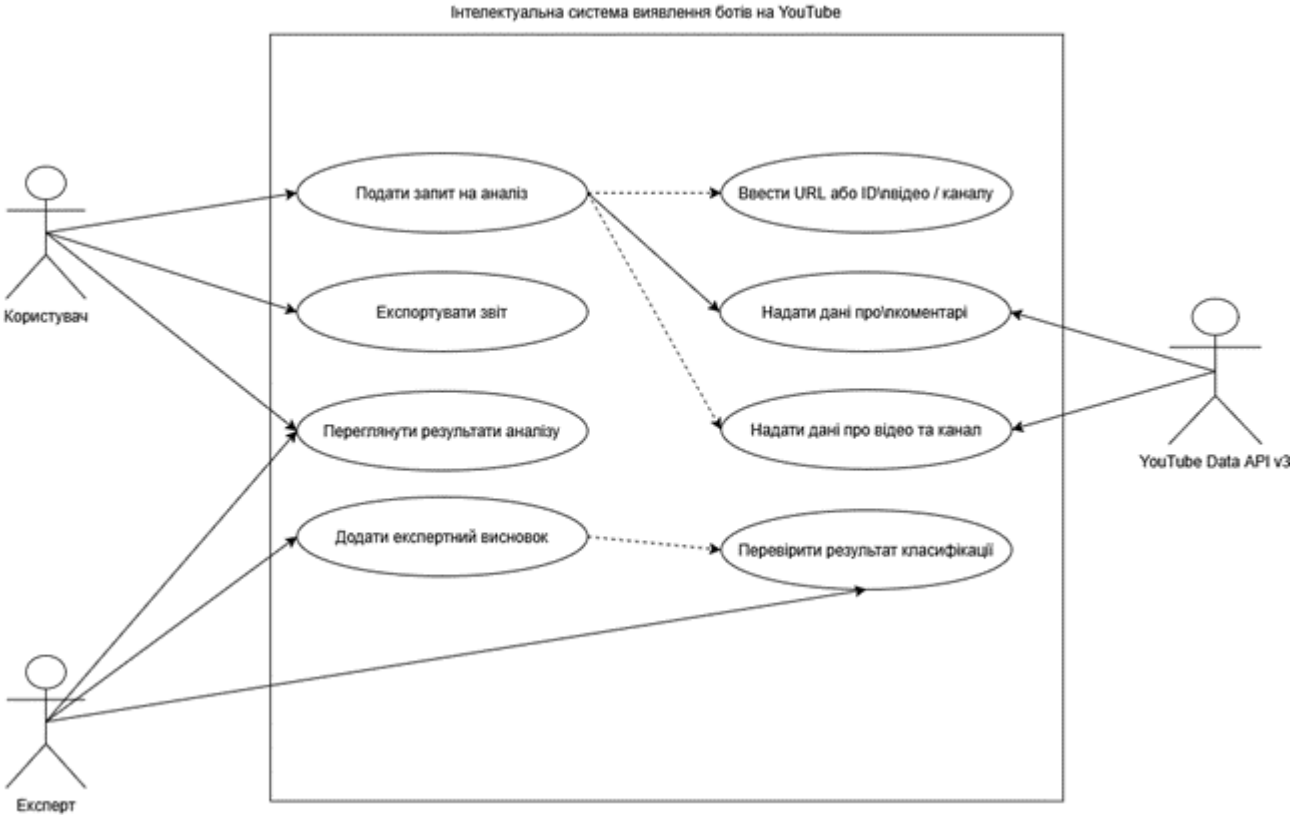
13. Martin Ester Hans-Peter Kriegel Jiirg Sander Xiaowei Xu. Density-Based algorithm for discovering clusters in large spatial databaseswith noise. URL: https://cdn.aaai.org/KDD/1996/KDD96-037.pdf?source=post_page----------.
14. Pca. *scikit-learn*. URL: https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html.
15. PostgreSQL: documentation. *PostgreSQL: The world's most advanced open source database*. URL: https://www.postgresql.org/docs/.
16. Python 3.13 documentation. *Python documentation*. URL: https://docs.python.org/3.13/.
17. TfidfVectorizer. *scikit-learn*. URL: https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html.
18. YouTube data API | google for developers. *Google for Developers*. URL: https://developers.google.com/youtube/v3?hl=ru.
19. Phenomenological model of information operation in social networking services. Molodetska K., Tymonin Y., Markovets O., Melnychyn A. Indonesian Journal of Electrical Engineering and Computer Science. 2020. Vol. 19, No. 2. PP. 1171–1180.
20. Maksim Kalameyets. Algorithms and techniques for bot detection in social networks. Library and information sciences. Université Paul Sabatier - Toulouse III; ITMO University, 2021
21. Hayawi K., Saha S., Masud M.M. et al. Social media bot detection with deep learning methods: a systematic review. Neural Comput & Applic, 2023, 35, P. 8903–8918.
22. Гончар І. А. Система визначення ботів на основі цифрового відбитку пристрою / І. А. Гончар. — 2021. — URL:https://ela.kpi.ua/server/api/core/bitstreams/10ab1602-aac7-44a9-8334-3040c9899b9a/content.
23. Люшенко Л., Перегуда Я. Спосіб побудови програмних детекторів для виявлення програмних ботів в соціальних мережах // Інформаційні

технології та суспільство. 2024. – Т. 1, №12. – С. 56-64. – URL: https://journals.maup.com.ua/index.php/it/article/view/3149/3592.

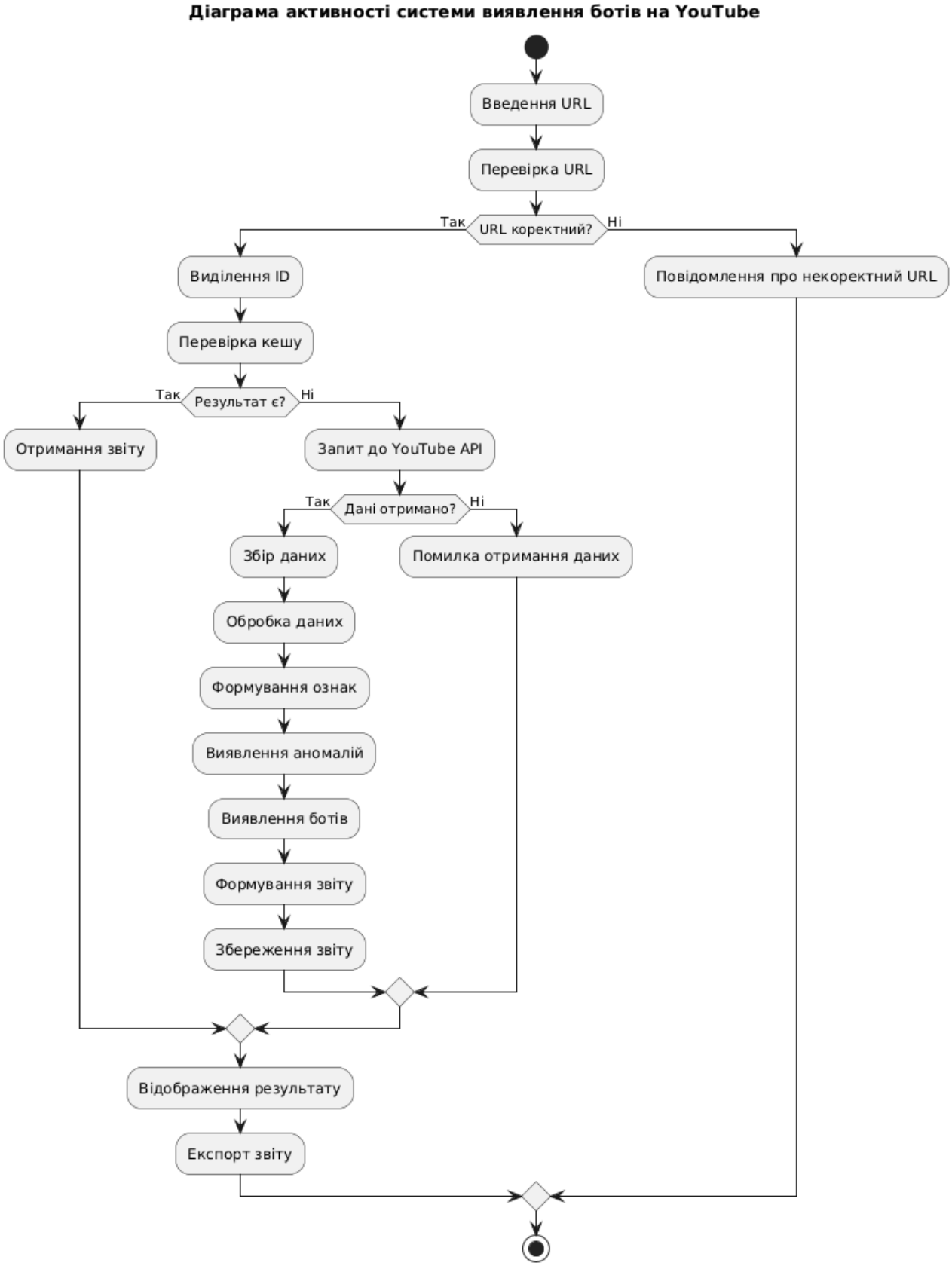
24. Pandas documentation – pandas 3.0.3 documentation. *pandas - Python Data Analysis Library*. URL: https://pandas.pydata.org/docs/whatsnew/v3.0.3.html
25. DBSCAN. scikit-learn 1.6.0 documentation. URL: https://scikit-learn.org/stable/modules/generated/sklearn.cluster.DBSCAN.html.
26. XGBoost Documentation – xgboost 3.3.0 documentation. *XGBoost Documentation – xgboost 3.3.0 documentation*. URL: https://xgboost.readthedocs.io/ .
27. Botometer X. *Botometer X*. URL: https://botometer.osome.iu.edu/
28. Julia Bohutska. Anomaly Detection – How to Tell Good Performance from Bad. URL: https://towardsdatascience.com/anomaly-detection-how-to-tell-good-performance-from-bad-b57116d71a10/ .
29. Ломберг Е. В. Виявлення координованої активності ботів у коментарях YouTube на основі аналізу текстових, часових і поведінкових ознак: **збірник праць учасників Всеукраїнської науково-практичної конференції здобувачів вищої освіти і молодих вчених**, м. Житомир, 22 травня 2026 р. Житомир : Поліський національний університет, 2026. С. 59–60.
30. Ломберг Е. В. Вплив соціальних ботів на сучасний інформаційний простір. **New Horizons in Scientific Research: Challenges and Solutions** : матеріали наукової конференції, 15–17 червня 2026 р С. 215–216.

ДОДАТКИ

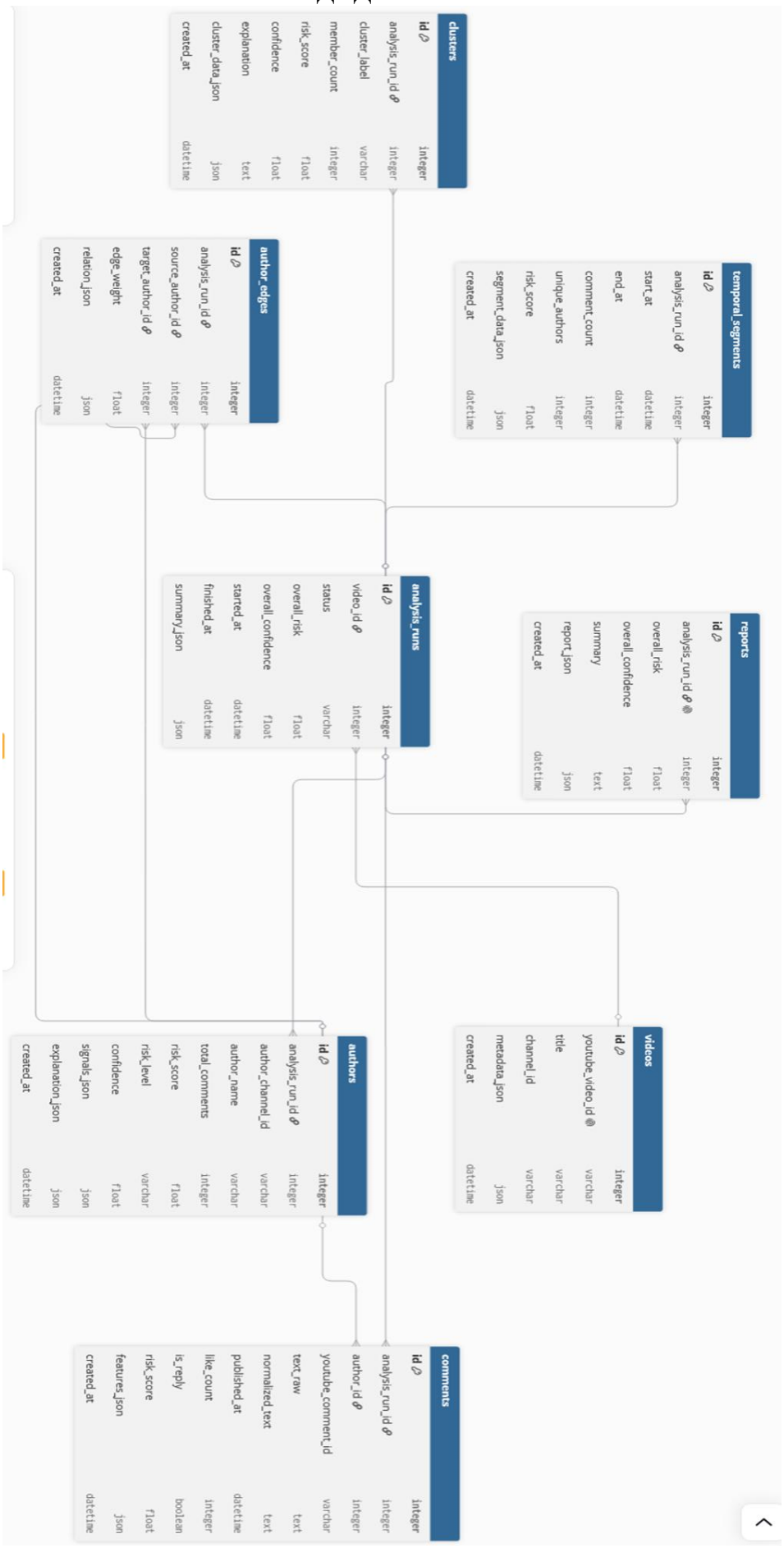
ДОДАТОК А



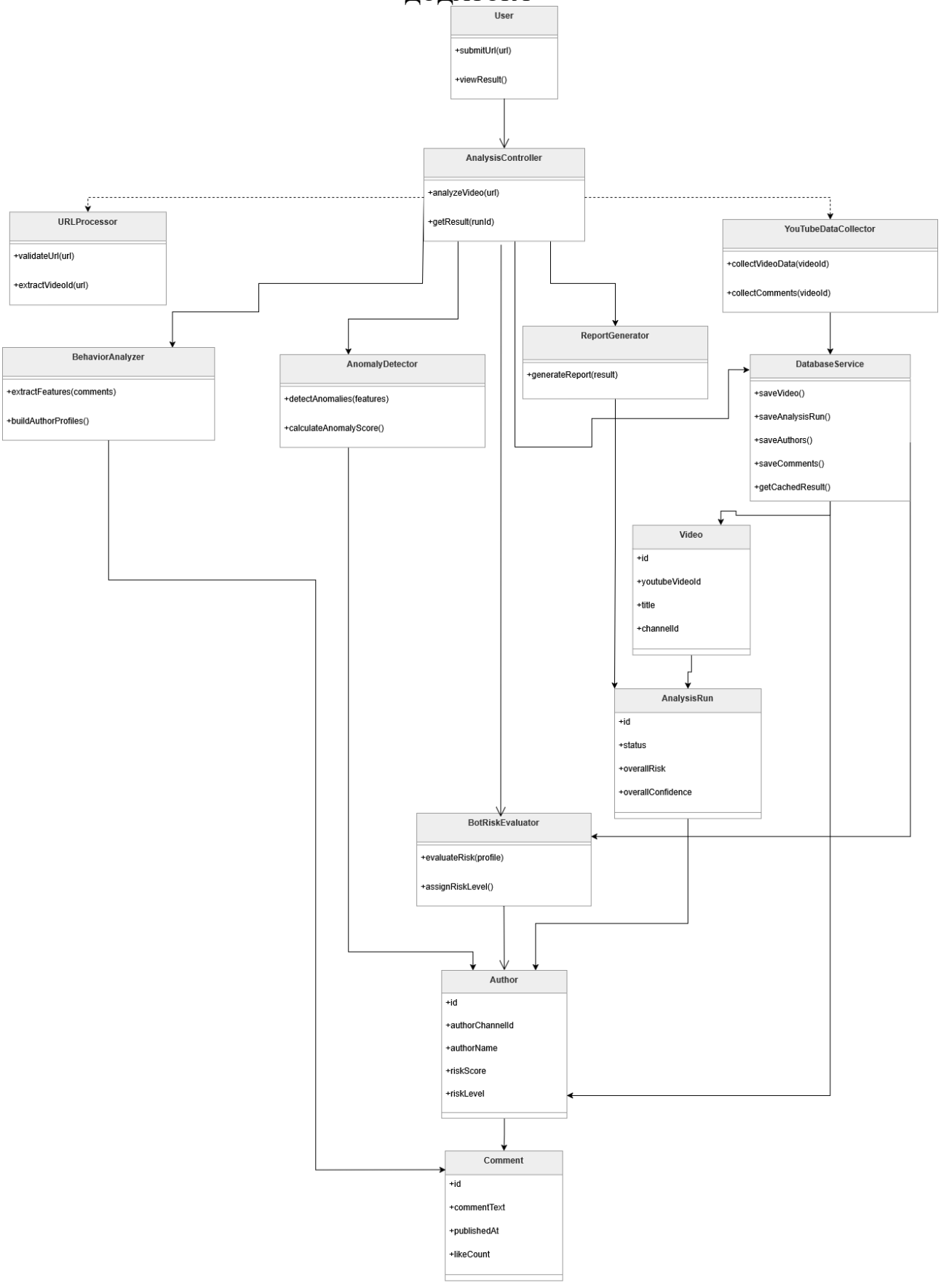
ДОДАТОК Б


Діаграма активності системи виявлення ботів на YouTube
Введення URL
Перевірка URL
URL коректний?
Так
Ні
Повідомлення про некоректний URL
Виділення ID
Перевірка кешу
Результат є?
Так
Ні
Отримання звіту
Запит до YouTube API
Дані отримано?
Так
Ні
Помилка отримання даних
Збір даних
Обробка даних
Формування ознак
Виявлення аномалій
Виявлення ботів
Формування звіту
Збереження звіту
Відображення результату
Експорт звіту

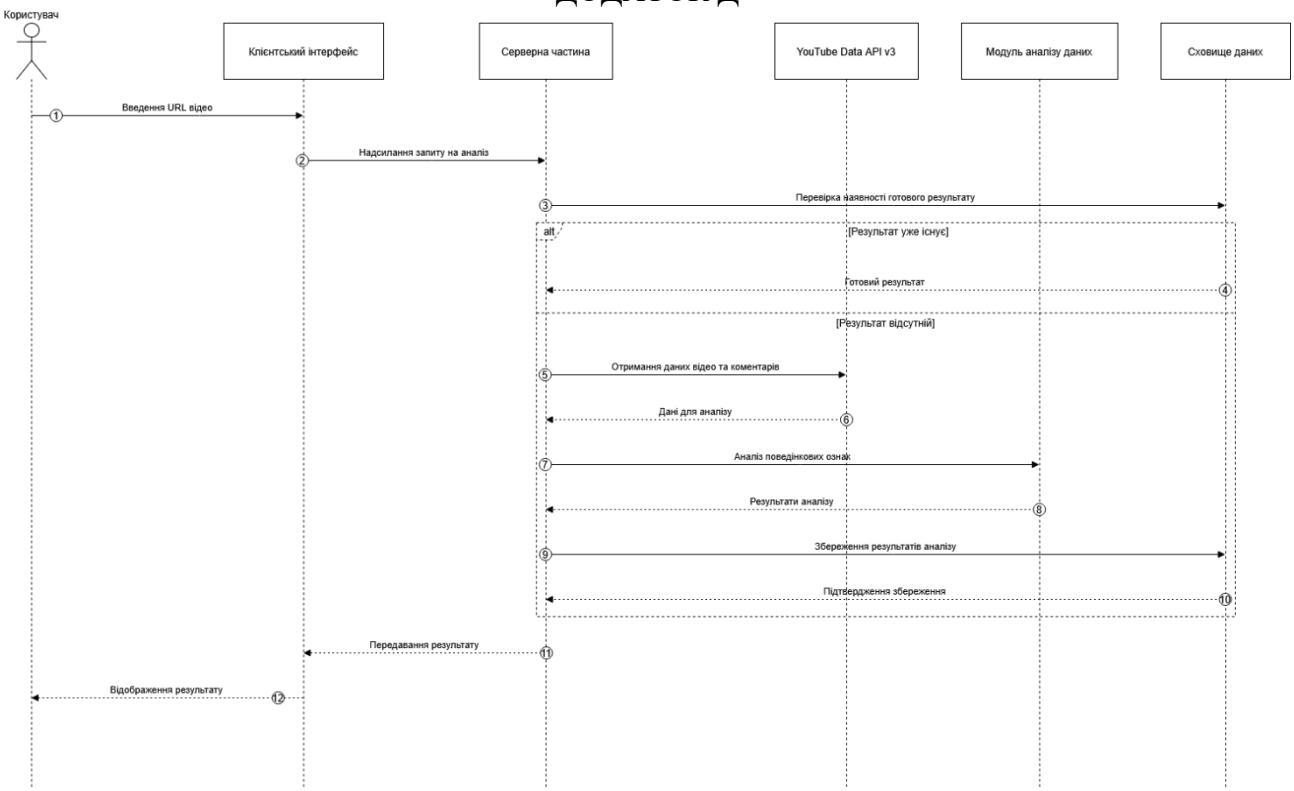
ДОДАТОК В


clusters
id integer
analysis_run_id integer
cluster_label varchar
member_count integer
risk_score float
confidence float
explanation text
cluster_data_json json
created_at datetime
temporal_segments
id integer
analysis_run_id integer
start_at datetime
end_at datetime
comment_count integer
unique_authors integer
risk_score float
segment_data_json json
created_at datetime
reports
id integer
analysis_run_id integer
overall_risk float
overall_confidence float
summary text
report_json json
created_at datetime
videos
id integer
youtube_video_id varchar
title varchar
channel_id varchar
metadata_json json
created_at datetime
author_edges
id integer
analysis_run_id integer
source_author_id integer
target_author_id integer
edge_weight float
relation_json json
created_at datetime
analysis_runs
id integer
video_id integer
status varchar
overall_risk float
overall_confidence float
started_at datetime
finished_at datetime
summary_json json
authors
id integer
analysis_run_id integer
author_channel_id varchar
author_name varchar
total_comments integer
risk_score float
risk_level varchar
confidence float
signals_json json
explanation_json json
created_at datetime
comments
id integer
analysis_run_id integer
author_id integer
youtube_comment_id varchar
text_raw text
normalized_text text
published_at datetime
like_count integer
is_reply boolean
risk_score float
features_json json
created_at datetime

ДОДАТОК Г


User
+submitUrl(url)
+viewResult()
AnalysisController
+analyzeVideo(url)
+getResult(runId)
URLProcessor
+validateUrl(url)
+extractVideoId(url)
YouTubeDataCollector
+collectVideoData(videoId)
+collectComments(videoId)
BehaviorAnalyzer
+extractFeatures(comments)
+buildAuthorProfiles()
AnomalyDetector
+detectAnomalies(features)
+calculateAnomalyScore()
ReportGenerator
+generateReport(result)
DatabaseService
+saveVideo()
+saveAnalysisRun()
+saveAuthors()
+saveComments()
+getCachedResult()
Video
+id
+youtubeVideoId
+title
+channelId
AnalysisRun
+id
+status
+overallRisk
+overallConfidence
BotRiskEvaluator
+evaluateRisk(profile)
+assignRiskLevel()
Author
+id
+authorChannelId
+authorName
+riskScore
+riskLevel
Comment
+id
+commentText
+publishedAt
+likeCount

ДОДАТОК Д


Користувач
Клієнтський інтерфейс
Серверна частина
YouTube Data API v3
Модуль аналізу даних
Сховище даних
1 Введення URL відео
2 Надсилання запиту на аналіз
3 Перевірка наявності готового результату
alt
[Результат уже існує]
Готовий результат
4
[Результат відсутній]
5 Отримання даних відео та коментарів
6 Дані для аналізу
7 Аналіз поведінкових ознак
8 Результати аналізу
9 Збереження результатів аналізу
10 Підтвердження збереження
11 Передавання результату
12 Відображення результату