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

Адамович Олексій Юрійович

УДК 004.738.5:316

КВАЛІФІКАЦІЙНА РОБОТА

Метод виявлення інформаційного впливу на користувачів в соціальних інтернет сервісах

125 - Кібербезпека

Подається на здобуття освітнього ступеня магістр

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

Керівник роботи
Веретюк Сергій Михайлович
старший викладач кафедри, к.т.н.

Житомир – 2023

АНОТАЦІЯ

Адамович О. Ю. Метод виявлення інформаційного впливу на користувачів в соціальних інтернет сервісах – Кваліфікаційна робота на правах рукопису.

Кваліфікаційна робота на здобуття освітнього ступеня магістр за спеціальністю 125 – кібербезпека. – Поліський національний університет, Житомир, 2023.

Анотація

Магістерська робота зосереджена на розробці інноваційного методу виявлення деструктивного впливу в соціальному інтернет-сервісі. В ході дослідження був проведений глибокий аналіз різноманітних типів деструктивних впливів, які можуть здійснюватись в онлайн-середовищі, а також вивчено існуючі методи їх виявлення. Оцінено переваги та недоліки цих методів з метою підвищення ефективності та точності виявлення негативного впливу.

В основі дослідження лежить застосування відкритих даних, зокрема YouTube API v3, для створення новаторського методу виявлення ознак деструктивного впливу. Розроблений метод ґрунтується на використанні порівняльного аналізу патернів "нормальної" та аномальної поведінки користувачів у коментарях до конкретного відео. Це дозволяє виявляти та класифікувати негативні впливи, сприяючи тим самим забезпеченню безпеки та позитивного інтерактивного середовища для користувачів соціального інтернет-сервісу.

Отримані результати роботи мають великий практичний потенціал і можуть знайти застосування в сферах соціальних мереж та інтерактивних платформ для ефективної боротьби з деструктивним впливом та забезпечення безпеки онлайн-спільнот.

Ключові слова: соціальний інтрнет сервіс, деструктивний влив, виявлення деструктивного впливу, виявлення аномалій.

ABSTRACT

Adamovych O. Yu. The method of detecting informational influence on users in social Internet services – Qualification work on the rights of the manuscript.

Qualification work for obtaining a master's degree in the specialty 125 - cyber security. – Polis National University, Zhytomyr, 2023.

Abstract

The master's thesis is focused on the development of an innovative method of detecting destructive influence in a social Internet service. In the course of the study, an in-depth analysis of various types of destructive influences that can be carried out in the online environment was carried out, as well as the existing methods of their detection were studied. The advantages and disadvantages of these methods were evaluated in order to increase the efficiency and accuracy of detecting negative effects.

The research is based on the use of open data, in particular the YouTube API v3, to create an innovative method for detecting signs of destructive influence. The developed method is based on the use of comparative analysis of patterns of "normal" and abnormal behavior of users in comments to a specific video. This allows identifying and classifying negative impacts, thereby contributing to the provision of safety and a positive interactive environment for users of the social Internet service.

The obtained results of the work have great practical potential and can be applied in the fields of social networks and interactive platforms to effectively combat destructive influence and ensure the safety of online communities.

Keywords: social internet service, destructive influence, detection of influence, detection of anomalies.

ЗМІСТ

СКОРОЧЕННЯ ТА УМОВНІ ПОЗНАЧЕННЯ

ІКС - Інформаційно-комунікаційна система

СІС – Соціальний інтрнет-сервіс

КТіМС – комп’ютерні технології і моделювання систем

ТЗ – технічні засоби

ОТЗ – основні технічні засоби

ІС – інформаційні системи

СВВ – системи виявлення вторгнень

СЗВ – системи запобігання вторгненням

ПІБ – політика інформаційної безпеки

КСЗІ – комплексна система захисту інформації

ВСТУП

Актуальність роботи. Актуальність теми розробки нових методів виявлення деструктивних впливів в соціальних інтернет-сервісах на сьогоднішній день надзвичайно важлива, оскільки зростає кількість випадків негативного впливу на користувачів цих платформ. Здатність швидко виявляти та ефективно протидіяти спробам дезінформації, кібербулінгу та інших шкідливих явищ в мережах сприятиме збереженню позитивного середовища для користувачів. Нові методи детекції також можуть відігравати ключову роль у попередженні маніпуляцій та впливу на громадську думку, що стає дедалі актуальнішим у сучасному інформаційному просторі. Розвиток і вдосконалення інструментів виявлення деструктивних впливів в соціальних інтернет-сервісах є стратегічно важливим завданням для забезпечення безпеки та етичності в цьому цифровому середовищі.

Метою роботи є розроблення методу виявлення деструктивного впливу на користувачів соціальних інтрнет сервісів на основі відкритих даних медіаплатформи YouTube.

Об’єктом дослідження деструктивні впливи в соціальних інтернет сервісах та їх ознаки.

Предмет дослідження є процес виявлення деструктивного впливу.

Методи дослідження: аналіз наукової літератури, статистичний аналіз, математичний аналіз, системний аналіз, теорія прийняття рішень.

Наукова новизна роботи: Розроблено метод виявлення деструктивних впливів в соціальних інтернет сервісах на основі відкритих даних.

Практичне значення отриманих результатів: Застосування розробленого методу дозволить підвищити ефективність систем аналізу та виявлення деструктивних впливів, в тому числі спрямованих на поширення дезінформації та просування деструктивних наративів.

Перелік публікацій за темою дослідження:

1. АдамовичО.Ю., Веретюк С. М. ГРАФОВІ ПІДХОДИ ДО МОДЕЛЮВАННЯ ТА АНАЛІЗУ СПІЛЬНОТ В ІНТЕРНЕТІ: РОЛЬ ТА

ЗАСТОСУВАННЯ СОЦІАЛЬНОЇ МЕРЕЖНОЇ АНАЛІТИКИ (SNA). ТРЕНДИ ТА ПЕРСПЕКТИВИ РОЗВИТКУ МУЛЬТИДИСЦИПЛІНАРНИХ ДОСЛІДЖЕНь : зб. матеріалів доп. учасн. IV Міжнародна студентська наукова конференція. м. Луцьк:, 2023. С. 294-297.

2. АдамовичО.Ю., Веретюк С. М. СУТНІСТЬ ТА ОСНОВНІ АСПЕКТИ ДЕЗІНФОРМАЦІЇ В СОЦІАЛЬНИХ МЕРЕЖАХ. Концепт науки XXI: стратегії, методи та наукові інструментиь : зб. матеріалів доп. учасн. IV Міжнародна студентська наукова конференція. м. Вінниця, 2023. С. 179-182.
3. АдамовичО.Ю., Веретюк С. М. ТЕХНОЛОГІЇ ВИЯВЛЕННЯ ДЕЗІНФОРМАЦІЇ В МЕРЕЖІ. «Інновації та науковий потенціал світу»: зб. матеріалів доп. учасн. III Міжнародна наукова конференція. м. Хмельницький, 2023. С. 169-170

РОЗДІЛ 1. АНАЛІЗ ОСОБЛИВОСТЕЙ ВИЯВЛЕННЯ ТА ЗАПОБІГАННЯ ПОШИРЕННЮ ДЕЗІНФОРМАЦІЇ В СОЦІАЛЬНИХ ІНТЕРНЕТ СЕРВІСАХ

1.1. Аналіз деструктивних впливів соціальних інтернет сервісах

Деструктивні впливи в соціальних мережах пройшли значний шлях еволюції від часів перших соціальних мереж до сучасних платформ. Історично, перші соціальні мережі, такі як інтернет-форуми та списки розсилки електронної пошти, слугували майданчиками для обміну інформацією та думками серед користувачів. Однак з появою більш складних соціальних мереж, таких як Facebook, Twitter, Instagram та інші, деструктивні впливи також стали складнішими та поширенішими.

Новим явищем є поширення дезінформації та фейкових новин через соціальні мережі. За допомогою масового розповсюдження неперевірених або навмисно спотворених даних через соціальні мережі, зловмисники можуть впливати на громадську думку, змінювати ставлення користувачів та створювати хаос у суспільстві. Такий деструктивний вплив може мати серйозні наслідки, включаючи розпалювання конфліктів, порушення довіри до засобів масової інформації та загрозу національній безпеці.

За останній час, спостерігається зростання кількості деструктивних впливів в соціальних мережах. Згідно з останніми дослідженнями, проведеними у 2023 році, обсяг фейкових новин і дезінформації на популярних соціальних платформах збільшився на 25% порівняно з попереднім роком. Це свідчить про серйозну та актуальну загрозу інформаційній безпеці та громадському порядку.

1.2. Аналіз засобів та каналів поширення дезінформації

Дезінформація в сучасному світі поширюється через різноманітні засоби та канали, які можна класифікувати за формами та способами подачі інформації. Ось

огляд основних засобів та каналів, які використовуються для поширення дезінформації:

- **Текстова дезінформація:**

Дезінформація у формі тексту часто приймає форму фейкових новин, які створюються з метою обману аудиторії. Це можуть бути вигадані події, спекуляції, або подача подій у специфічний спосіб для досягнення певного ефекту. Вмкористання недостовірних статистичних даних або перекручування фактів для створення ілюзії підтвердження своїх тверджень.

- **Відеодезінформація:**

Це підроблені аудіовізуальні записи, створені за допомогою штучного інтелекту, в яких образи, слова або сцени можуть бути змінені для створення хибної інформації. Діпфейки можуть бути використані для дискредитації осіб або створення хибного враження про події. Використання програмного забезпечення для монтажу та обробки відео дозволяє змінювати зміст та контекст відеозаписів для створення обманливих матеріалів.

- **Аудіальна дезінформація:**

Дезінформатори створюють фальшиві аудіозаписи, включаючи вигадані інтерв'ю, інсценовані події або маніпульовані розмови, які можуть бути поширені через подкасти та інші аудіальні платформи.

- **Соціальні мережі та онлайн-платформи:**

Популярні соціальні мережі, такі як Facebook, Twitter, Instagram, стали важливими каналами для поширення дезінформації. Дезінформатори використовують ці платформи для розповсюдження фейкових новин, фальшивих фотографій та відео. Месенджери, такі як WhatsApp, Telegram, стали популярними серед дезінформаторів для поширення дезінформації у вигляді текстових повідомлень, фото і відео. Сервіси, такі як YouTube, Vimeo, дозволяють дезінформаторам завантажувати та поширювати відео.

- **Координована неавтентична поведінка:**

Для поширення дезінформації створюються фейкові акаунти на соціальних мережах та інших платформах. Ці акаунти можуть бути використані для масового ретвіту, шерингу, лайків або коментарів, що підсилюють поширення дезінформації.

- **Класичні медіа та друковані видання:**

Деякі дезінформатори використовують класичні медіа та друковані видання для поширення дезінформації, подаючи її як авторитетні новини.

Важливо враховувати, що поширення дезінформації може мати серйозні наслідки для суспільства, політичних процесів та громадського довір'я. Боротьба з нею вимагає розвинених методів виявлення та попередження, а також медіаграмотності для сприйняття інформації критично та обачно.

1.3. Аналіз методів поширення дезінформації

Методи поширення дезінформації є різноманітними і вдосконалюються з розвитком технологій та доступу до медіа. Фейкові новини стали основним інструментом впливу на громадську думку. Вони можуть виникати як на основі створення вигаданих історій, так і на перекручуванні реальних подій. Використання незавідомих джерел та анонімних авторів у новинах робить їх важкими до перевірки для звичайного читача.

Діпфейки, зокрема відеодіпфейки, стають все більш аутентичними завдяки розвитку штучного інтелекту та глибокого навчання. Ця технологія може створити переконливі відеозаписи, які навіть професіонали можуть важко відрізнити від реальних. Це підвищує загрозу використання діпфейків у політичних кампаніях та інших сферах.

Соціальні мережі та месенджери стають основними каналами поширення дезінформації через їхню широку аудиторію та можливість швидкого ретвіту та репостів. Створення фейкових аккаунтів та ботів для автоматизації поширення дезінформації робить цей процес ще більш масштабним та важким до виявлення.

Фальшиві веб-сайти стають дедалі більш вдосконаленими у своєму дизайні та вмінні створювати вигляд авторитетних джерел. Це робить їхню роботу більш

ефективною в обмані читачів. Координована неавтентична поведінка, зокрема за допомогою масових акаунтів із схожими тезами, створює ілюзію підтримки та популярності дезінформаційного контенту.

1.4. Аналіз методів протидії поширення дезінформації

Методи протидії дезінформації можна розділити на короткострокові і довгострокові за тривалістю їхнього впливу. Довгострокові стратегії виявляються особливо важливими, оскільки вони спрямовані на зменшення загальної вразливості до дезінформації у майбутньому. З точки зору часового аспекту, існують превентивні методи, спрямовані на конкретне питання та передбачають співпрацю на регіональному рівні. Також існують негайні заходи, включаючи кризову комунікацію та фактчекінг, які реагують негайно на поширення дезінформації та коригують неправильну інформацію. Довгострокові стратегії включають в себе заходи, спрямовані на підвищення медіаграмотності громадян, формування соціальних норм щодо відповідального споживання інформації та співпрацю між різними стейкхолдерами. Такий підхід дозволяє ефективно враховувати як негайну потребу в реакції на дезінформацію, так і необхідність створення стійких механізмів для запобігання її поширенню у майбутньому.

Технічні методи протидії поширенню дезінформації в кіберпросторі відіграють важливу роль у забезпеченні інформаційної безпеки та нейтралізації дезінформаційних загроз. Однією з ключових стратегій є аналіз великих обсягів даних. Використання аналітичних інструментів та інформаційних технологій дозволяє автоматично виявляти патерни та аномалії в потоці інформації. Машинне навчання та штучний інтелект можуть аналізувати поведінку користувачів та ідентифікувати можливі спроби маніпуляції аудиторією.

Далі, технічні методи включають в себе виявлення фейків і діпфейків. Спеціалізовані алгоритми можуть аналізувати текстовий та візуальний контент, виявляти недостовірність інформації та підробки. Використання комп'ютерного

зору допомагає перевіряти відео та аудіозаписи на наявність ознак монтажу та обробки. Додатково, технічні методи включають моніторинг соціальних мереж та онлайн-платформ. Автоматизовані системи можуть постійно відслідковувати активність на цих платформах та ідентифікувати джерела поширення дезінформації. Також, технічні фільтри можуть застосовуватися для фільтрації шкідливого контенту та блокування облікових записів, що активно розповсюджують недостовірну інформацію.

Не менш важливим аспектом є кіберзахист інфраструктури. Заходи кіберзахисту спрямовані на запобігання несанкціонованому доступу до інформації та захист від можливих атак, спрямованих на поширення дезінформації. Ці заходи включають в себе криптографічний захист, захист мережевих інфраструктур та безпеку даних.

Усі ці технічні методи разом сприяють створенню стійкого інформаційного середовища, де інформація є надійною та об'єктивною. Вони дозволяють ефективно виявляти та нейтралізувати загрози дезінформації, забезпечуючи надійний кіберзахист та інформаційну безпеку.

Поза технічними методами, існують і соціокультурні стратегії для протидії поширенню дезінформації. Ці стратегії відіграють важливу роль у формуванні стійкого інформаційного середовища та впливають на спосіб, яким індивіди сприймають та реагують на інформацію.

Медіаграмотність та освіта є однією з ключових стратегій. Програми медіаграмотності допомагають людям розрізняти надійну інформацію від дезінформації. Соціокультурні стратегії також включають встановлення соціальних норм та цінностей, які не сприяють поширенню дезінформації. Підтримка ініціатив, що визнають важливість об'єктивності та достовірності інформації, може стимулювати громадян до відповідальної споживчої поведінки.

Співпраця та колаборації грають також важливу роль. Встановлення механізмів співпраці між громадськими організаціями, медіа та урядовими інституціями дозволяє виявляти та нейтралізувати дезінформаційні загрози шляхом обміну інформацією та ресурсами.

Іншим важливим аспектом є забезпечення інформаційної прозорості і відкритості. Урядові та корпоративні структури повинні сприяти відкритості та доступності інформації для громадськості, що сприятиме підвищенню довіри та запобіганню розповсюдженню дезінформації.

Отже, розробка та впровадження ефективних правових механізмів для боротьби з поширенням дезінформації є важливою складовою соціокультурних стратегій. Законодавство щодо відповідальності за розповсюдження недостовірної інформації може стимулювати відповідальну поведінку та обмежувати дії дезінформаторів.

Ці соціокультурні стратегії, разом із технічними методами, формують комплексний підхід до протидії дезінформації. Вони сприяють створенню стійкого інформаційного середовища, де інформація є надійною та об'єктивною, а громадяни мають навички та засоби для захисту від маніпуляцій та обману.

Висновки до першого розділу

1. Проведено аналіз деструктивних впливів соціальних інтернет-сервісів та методів поширення дезінформації. Показано, що деструктивні впливи в соціальних мережах пройшли значний шлях еволюції, зокрема завдяки розвитку більш складних соціальних мереж. Одним з актуальних явищ є поширення дезінформації та фейкових новин через соціальні мережі, що створює загрозу для громадської думки, довіри до мас-медіа та національної безпеки.
2. Розглянуто різні засоби та канали поширення дезінформації, такі як текстова, відео- та аудіальна дезінформація, а також використання соціальних мереж, месенджерів, відео-хибних записів та класичних медіа для цілей дезінформації. Не менш важливим є аналіз методів поширення дезінформації, включаючи використання фейкових новин, діпфейків, соціальних мереж та координованої неавтентичної поведінки.
3. Для подолання цієї серйозної загрози інформаційній безпеці та громадському порядку необхідні розвинені методи виявлення та попередження дезінформації, а

також підвищена медіаграмотність громадян для критичного та обачного сприйняття інформації. Тому завдання удосконалення та розроблення ефективних методів виявлення деструктивних впливів в соціальних інтернет-сервісах є актуальною задачею.

РОЗДІЛ 2. РОЗРОБЛЕННЯ МЕТОДУ ВИЯВЛЕННЯ ІНФОРМАЦІЙНОГО ВПЛИВУ НА КОРИСТУВАЧІВ В СОЦІАЛЬНИХ ІНТЕРНЕТ СЕРВІСАХ

2.1 Особливості деструктивних інформаційних впливів в соціальних інтернет сервісах на прикладі медіаплатформи YouTube

За даними дослідження [13], у вересні 2023 році медіаплатформа YouTube займає друге місце у світі за популярністю серед соціальних інтернет-сервісів з кількістю активних користувачів (акторів) понад 2.7 млрд. Сучасні медіаплатформи складають значну конкуренцію традиційним засобам масової інформації - телебаченню, радіо, друковані виданням. Популярність медіаплатформ пояснюється перевагами технології WEB 2.0 та реалізацією концепцію “many-to-many” – в основі їх функціонування лежить принцип, коли користувачі соціального інтернет сервісу можуть бути не тільки споживачами медіа-контенту, а й його генераторами (авторами або власниками).

В контексті інформаційної безпеки основний ризик для користувачів медіаплатформи полягає в низькій ефективності засобів регулювання поведінки користувачів, не ефективними засобами валідації контенту, формальній модерації та цензуруванні форми, а не змісту [14]. Як наслідок, медіаплатформа Youtube перетворилися на дієвий інструмент проведення інформаційних операцій, кампаній із поширення дезінформації, інформаційно-психологічного впливу на акторів сервісів.

Таблиця 1. Класифікація впливів

	Назва впливу	***Дія***	***Потенційна шкода***
1	**Кібербулінг та цькування** (агресивні дії, що мають на меті заподіяти шкоду іншій людині через Інтернет)	образи, погрози, поширення фальшивих повідомлень тощо	виведення людини із психологічної рівноваги
2	**Експлуатація** (використання соціальних мереж для отримання персональної інформації про людей з метою її зловживання, викрадення, шантажу тощо)	фішинг, обман, направлення на фейкові платіжні системи тощо	втрата матеріальних цінностей
3	**Дезінформація та фейкові новини** (вплив на громадську думку або культивація паніки та хаосу)	розповсюдження неправдивої інформації або фальшивих новин	соціальний контроль та вплив на суспільство, провокації, елемент гібридної війни
4	**Залежність від соціальних мереж** (користувачі стають залежними від соціальних мереж і витрачають на них більше часу, ніж це вважається здоровим.)	нав'язливі ресурси, онлайн ігри, онлайн знайомства	десоціалізація (втрата контролю над життям, розірвання соціальних та

			культурних зв'язків)
5	**Кіберзлочинність** (використання соціальних мереж для вчинення злочинів)	крадіжка персональних даних, фішинг, несанкціонований доступ до інформаційних систем тощо	втрати матеріальних цінностей в великих масштабах
6	**Психологічна шкода** (погіршення психічного здоров'я, депресію, тривогу та інші психічні проблеми)	булінг, депресивна інформація, суїцідальні мотивації	депресія, виведення із психологічної рівноваги
7	**Втрата приватності** (приватне життя стає публічним, нанесення шкоди репутації, родині тощо)	незаконне збирання персональної інформації	втрата репутації, шантаж

2.2 Маркери інформаційного впливу в соціальному інтернет-сервісі YouTube та організація процесу збору даних

В основі методу виявлення інформаційного деструктивного впливу на віртуальну спільноту в соціальному інтернет сервісі Youtube лежить пошук аномалій на різних інформаційних рівнях, тобто ознак поведінки користувача або користувачів, які виділяються на фоні “нормального” патерну. Основні ознаки встановлено в [6, 7], запропонований в цій статті метод використовує:

- тональність повідомлення;
- динаміка повідомлень (потік повідомлень, коментарів на годину);
- час створення повідомлення (наприклад, невідповідність між часом публікації та появою коментарів);
- час реєстрації користувача.

Як було зазначено вище соціальні інтернет-сервіси надають відкритий доступ до деперсоніфікованої інформації через API. В роботі використано YouTube Api v3 [15]. Інформаційна схема розміщення медіа контенту на платформі YuoTube має ієрархічний характер (рис.2.1).

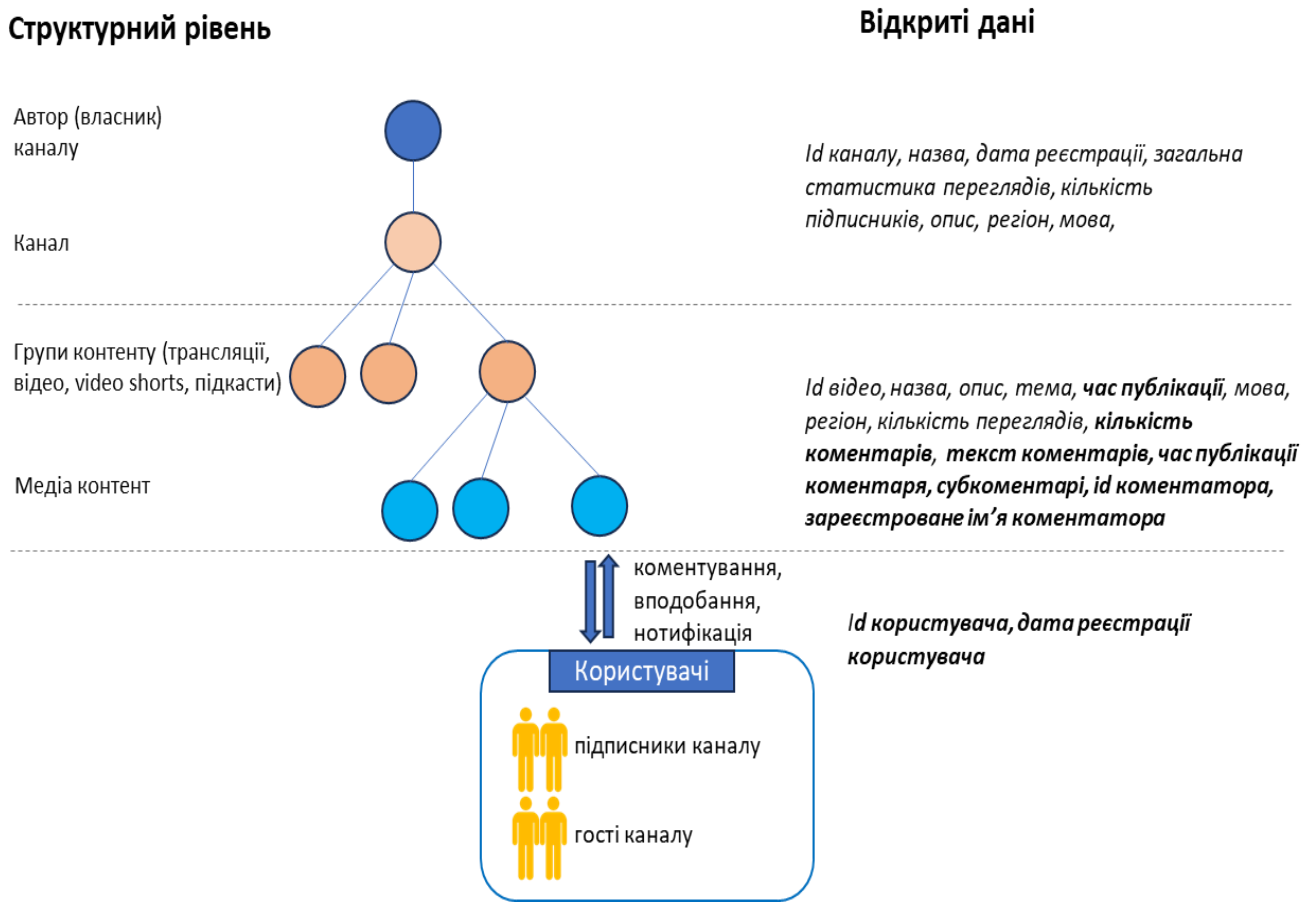


Рисунок 2.1. Інформаційна схема розміщення медіаконтенту та відповідні набори відкритих доступних даних в соціальному інтернет сервісі YouTube.

3.3 Метод виявлення деструктивного впливу на користувачів в соціальному інтернет сервісі YouTube.

Метод передбачає попередній аналіз не менше ніж 20 наборів даних для формування патернів, які віднесено до класу “нормальних”.

Крок 1. Збір та структурування метаданих для аналізу наявності деструктивного інформаційного впливу в межах простору коментарів під конкретним відео.

Крок 2. Формування часових рядів з інформацією кількість коментарів на годину, кількість коментарів за добу. Довжина вибірки складає 240 годин. $S_i = S(t_i)$

Крок 3. Нормалізація часових рядів за правилом:$Sn_i = \frac{S_i}{max(S(t_i)}$

Крок 4. Статистична систематизація отриманих часових рядів - формування розподілу коментарів за 24 години - Q_1

Крок 5. Формування статистичного розподілу на основі аналізу часу реєстрації користувачів - Q_2.

Крок 6. Порівняння нормалізованого часового ряду з патерном на основі розрахунку коефіцієнта кореляції $K_{\text{кор}}$, порівняння статистичних розподілів коментарів та часу реєстрації користувачів з нормальними патернами (на основі розрахунку відстані Кульбака-Лейблера) $D_{KL}(P_1||Q_1)$, $D_{KL}(P_2||Q_2)$, $Q_{1,2}$ - відповідно розподіли характеристик нормальних патернів.

Крок 7. Розрахунок інтегрального індексу та порівняння з пороговим значенням $I_{max} = 0.9$:

$$I = 1 - (\alpha \cdot (1 - K_{\text{кор}}) + \beta \cdot D_{KL}(P_1||Q_1) + \gamma \cdot D_{KL}\big(P_2\big|\big|Q_2\big)\big), \alpha = 0.2, \beta = 0.4, \gamma = 0.4$$

Крок 8. Якщо на кроці 7 не виявлено аномалію, метод передбачає подальший аналіз тональності. Наявні бібліотеки маркування тональності не забезпечують ефективну обробку української мови, то крок передбачає напівавтоматизовану обробку множини коментарів. Розмітку зроблено за трьома класами: негативно, нейтрально, позитивно.

Крок 9. За результатами кроку формується інтегральний часовий ряд $g(t_i)$, який характеризує динаміку тональності за час спостереження (240 годин). В подальшому визначається модуль суми значень функції, який порівнюється з пороговим значенням G_{MAX}. Структура методу представлено на рис.2.2

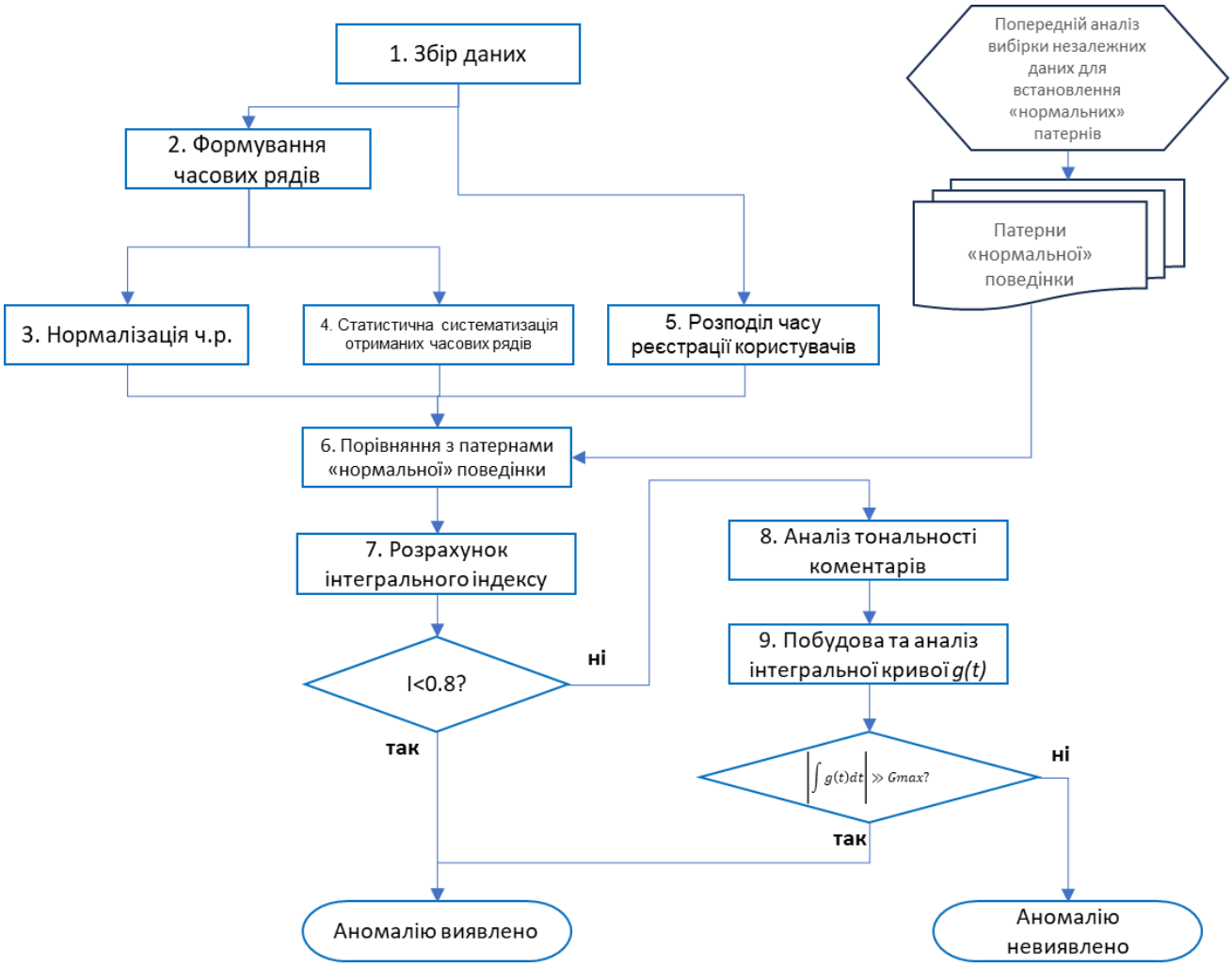


Рисунок 2.2 Алгоритм ідентифікації деструктивного інформаційного впливу

Висновки до другого розділу

Проаналізовано актуальні методи виявлення деструктивних інформаційних впливів на віртуальні спільноти в соціальних інтернет сервісах.

На основі аналізу деперсоніфікованих відкритих даних, які доступні в медіаплатформі Youtube, запропоновано метод виявлення штучного інформаційного впливу на віртуальну спільноту.

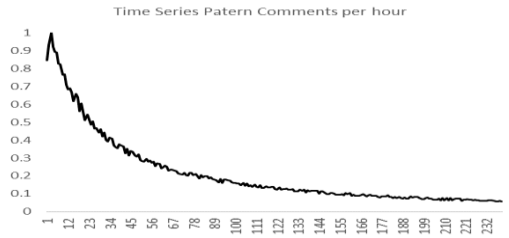
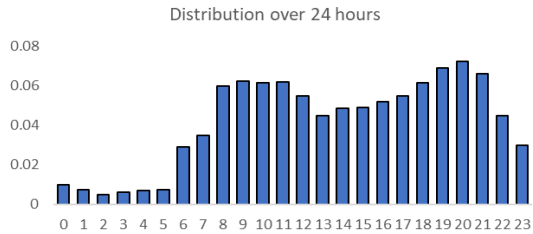
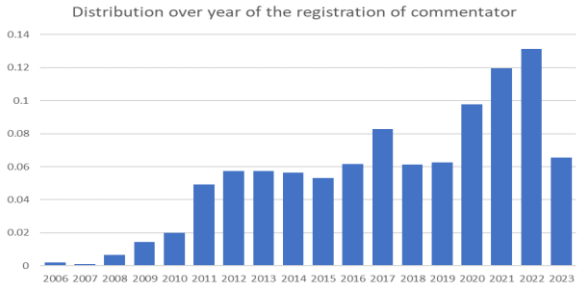
Метод передбачає виявлення аномальної поведінки учасників віртуальної спільноти і передбачає проведення порівняльного аналізу на різних аспектів поведінки користувачів – динаміка коментарів в часі, активності коментаторів протягом доби, час реєстрації користувача. Як додаткову опцію метод передбачає додатковий аналіз тональності коментарів.

РОЗДІЛ 3. АПРОБАЦІЯ МЕТОДУ ВИЯВЛЕННЯ ДЕСТРУКТИВНОГО ВПЛИВУ НА КОРИСТУВАЧІВ В СОЦІАЛЬНИХ ІНТЕРНЕТ СЕРВІСАХ

3.1 Побудова патернів нормальної поведінки користувачів в соціальному інтернет-сервісі YouTube

Для побудови патернів було обрано випадковим чином 20 відео з різних каналів Youtube. Мова -українська, кількість переглядів більше 25000, кількість коментарів не більше 2000, тематика - соціально-політична. Нормовані часові ряди, розподіл Q1 та Q2 наведено в табл.3.1

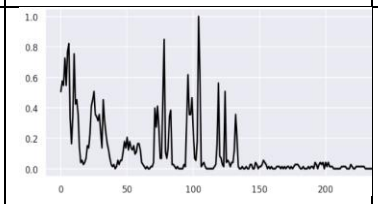
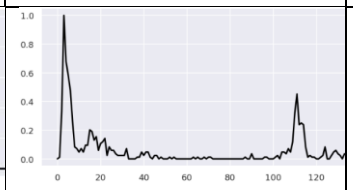
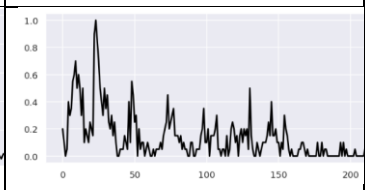
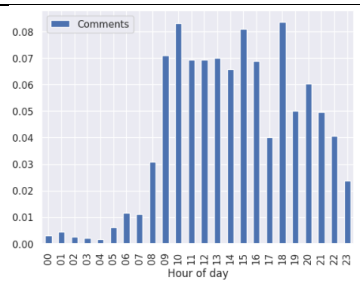
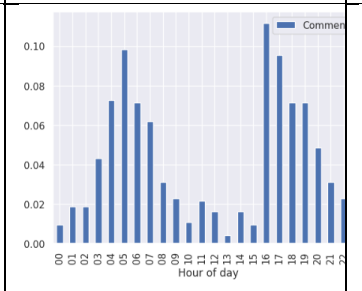
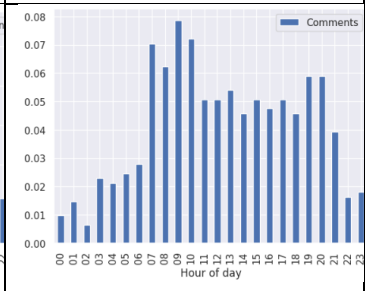
Таблиця 3.1 Показники «нормального» патерну

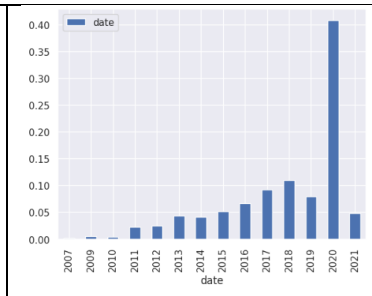
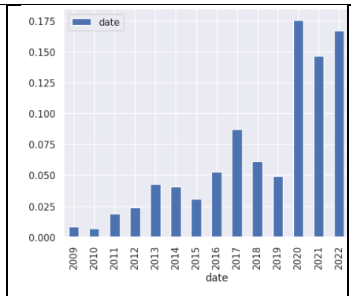
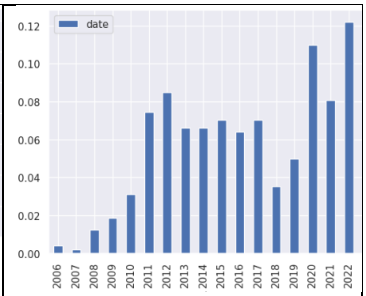
Нормалізований часовий ряд	
Розподіл коментарів по годинам доби, Q_1	
Розподіл часу реєстрації коментаторів,, Q_2	

3.2 Результати застосування методу виявлення деструктивного впливу в соціальному інтернет-сервісі.

Для ідентифікації інформаційного деструктивного впливу було обрано три відео на соціальну та політичну тематику, результати застосування методу виявлення інформаційного впливу на віртуальну спільноту в соціальному інтернет сервісі YouTube наведено в табл.3.2

Таблиця 3.2. Результати дослідження

Сценарій	**1**	**2**	**3**
id	cOhvA4JUDqE	LyAH6ma9MJA	gY4Agmj6ES
Дата публікації	05.03.2021	24.08.2023	17.06.2023
Кількість коментарів	1974	1205	1041
Кількість унікальних коментаторів	1316	586	643
Кількість переглядів	1.3 млн	140 тис.	82 тис.
Нормалізований часовий ряд			
Коефіцієнт кореляції	0.57	0.51	0.61
Розподіл коментарів по годинам доби, P_1			
$D_{KL}(P_1 \| Q_1)$	0.075	0.588	0.083

Розподіл часу реєстрації коментаторів, P_2			
$D_{KL}(P_2 \| Q_2)$	0.558	0.088	0.046
Інтегральний індекс	**0.66**	**0.63**	0.87
Висновок	Вплив наявний	Вплив наявний	Вплив не виявлено

Відповідно до запропонованого методу для Сценарію №3 необхідно провести додатковий аналіз тональності тексту коментарів. Для цього була реалізована експертна оцінка коментарів, виконано відповідну розмітку коментарів. Результати аналізу тональності, який визначає тон і емоційне ставлення коментаторів до досліджуваного відео, наведено на рис. 3.1

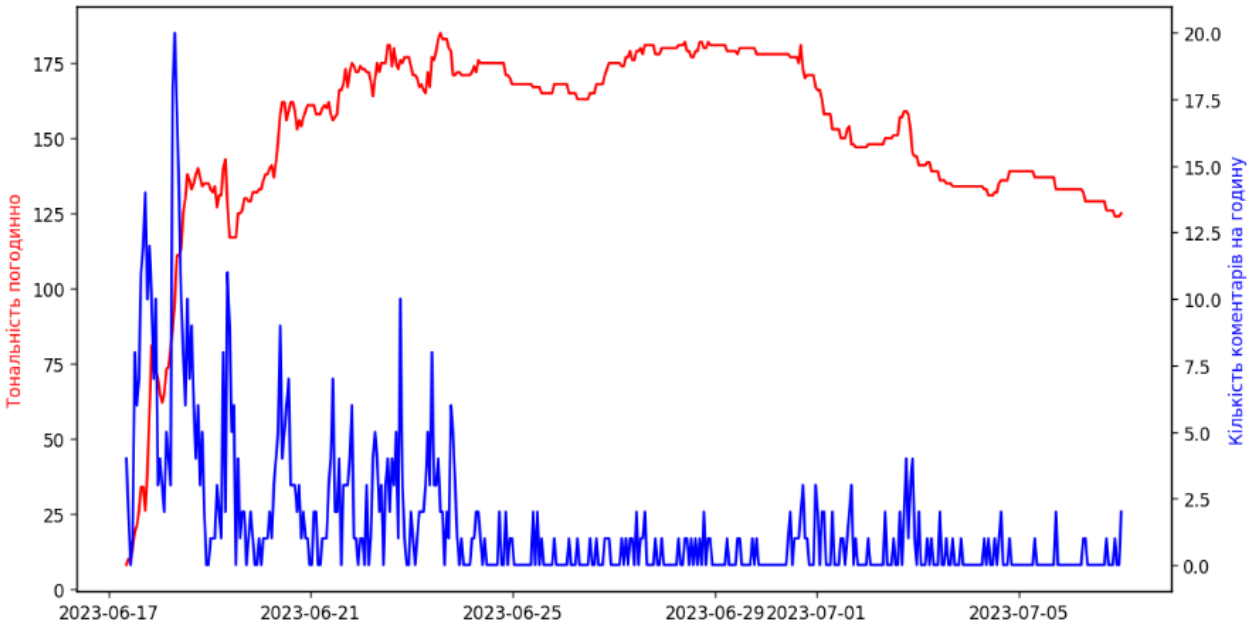


Рисунок 3. 1 Результати аналізу тональності для сценарію 3.

Висновки до третього розділу

1. Проаналізовано патерни «нормальної» поведінки користувачів - побудовані криві розподілу активності коментаторів
2. Наведено результати експерименту (застосування методу).
3. Встановлено, що запропонований метод дозволяє визначати наявність інформаційного (штучного) впливу на віртуальну спільноту.

ЗАГАЛЬНІ ВИСНОВКИ

1. Проведено аналіз деструктивних впливів соціальних інтернет-сервісів та методів поширення дезінформації. Показано, що деструктивні впливи в соціальних мережах пройшли значний шлях еволюції, зокрема завдяки розвитку більш складних соціальних мереж. Одним з актуальних явищ є поширення дезінформації та фейкових новин через соціальні мережі, що створює загрозу для громадської думки, довіри до мас-медіа та національної безпеки.
2. Розглянуто різні засоби та канали поширення дезінформації, такі як текстова, відео- та аудіальна дезінформація, а також використання соціальних мереж, месенджерів, відео-хибних записів та класичних медіа для цілей дезінформації. Не менш важливим є аналіз методів поширення дезінформації, включаючи використання фейкових новин, діпфейків, соціальних мереж та координованої неавтентичної поведінки.
3. На основі аналізу деперсоніфікованих відкритих даних, які доступні в медіаплатформі Youtube, запропоновано метод виявлення штучного інформаційного впливу на віртуальну спільноту. Метод передбачає виявлення аномальної поведінки учасників віртуальної спільноти і передбачає проведення порівняльного аналізу на різних аспектів поведінки користувачів – динаміка коментарів в часі, активності коментаторів протягом доби, час реєстрації користувача. Як додаткову опцію метод передбачає додатковий аналіз тональності коментарів.
4. Проведено апробацію запропонованого методу виявлення деструктивних впливів в соціальних інтернет-сервісах. Зокрема, проаналізовано патерни «нормальної» поведінки користувачів -побудовані криві розподілу активності коментаторів. Наведено результати експерименту (застосування методу). Встановлено, що запропонований метод дозволяє визначати наявність інформаційного (штучного) впливу на віртуальну спільноту.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Дубов Д. В., Ожеван М. А. Кібербезпека: світові тенденції та виклики для України. Аналітична доповідь. [Текст] – К. : НІСД, 2011. – 30 с.
2. Додонов О.Г., Горбачик О.С., Кузнєцова М.Г. Інформаційне суспільство: технології та безпека // Інформація та відкритість влади як засоби демократизації суспільства: Зб. матеріалів «круглого столу». [Текст] / К.: Альтпрес. – 2003. – С. 119-124.
3. Горбулін В.П., Качинський А.Б. Методологічні засади розробки стратегії національної безпеки // Стратегічна панорама. [Текст] / Горбулін В.П. – 2004. – № 3. – С. 15 - 24.
4. Молодецька-Гринчук, К. (2017). Метод оцінювання ознак загроз інформаційній безпеці держави у соціальних інтернет-сервісах. Автоматизация технологических и бизнес-процессов, (9, Iss. 2), 36-42.
5. Berghel Hal. Malice domestic: The Cambridge analytica dystopia. Computer. 2018. № 5. С. 84−89.
6. Mangold, W. G., & Faulds, D. J. (2009). Social media: The new hybrid element of the promotion mix. Business horizons, 52(4), 357-365.
7. Marathe, S., & Shirsat, K. P. (2015). Approaches for mining youtube videos metadata in cyber bullying detection. International Journal of Engineering Research & Technology, 4(5), 680-684.
8. Kirdemir, B., Adeliyi, O., & Agarwal, N. (2022, April). Towards characterizing coordinated inauthentic behaviors on YouTube. In ROMCIR 2022 CEUR Workshop Proceedings (Vol. 3138, pp. 100-116).
9. Sai, L. (2019). YouTube Video Bot Detection–A Deep Learning-Based Framework.
10. Yang, K. C., Varol, O., Davis, C. A., Ferrara, E., Flammini, A., & Menczer, F. (2019). Arming the public with artificial intelligence to counter social bots. Human Behavior and Emerging Technologies, 1(1), 48-61.

11. Obadimu, A., Khaund, T., Mead, E., Marcoux, T., & Agarwal, N. (2021). Developing a socio-computational approach to examine toxicity propagation and regulation in COVID-19 discourse on YouTube. Information Processing & Management, 58(5), 102660.
12. Chowdhury, S., Khanzadeh, M., Akula, R., Zhang, F., Zhang, S., Medal, H., ... & Bian, L. (2017). Botnet detection using graph-based feature clustering. Journal of Big Data, 4, 1-23.
13. Wang, J., & Paschalidis, I. C. (2014, September). Botnet detection using social graph analysis. In 2014 52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton) (pp. 393-400). IEEE.
14. https://www.demandsage.com/youtube-stats/
15. Trana, R.E., Gomez, C.E., Adler, R.F. (2021). Fighting Cyberbullying: An Analysis of Algorithms Used to Detect Harassing Text Found on YouTube. In: Ahram, T. (eds) Advances in Artificial Intelligence, Software and Systems Engineering. AHFE 2020. Advances in Intelligent Systems and Computing, vol 1213. Springer, Cham. https://doi.org/10.1007/978-3-030-51328-3_2
16. Youtube Data API | Google Developers, Google Developers, 2023, https://developers.google.com/youtube/v3.