

Міністерство освіти і науки України
Київський національний університет імені Тараса Шевченка
Навчально-науковий інститут філології
кафедра української мови та прикладної лінгвістики

**Автоматичне розпізнавання
Телеграм-каналів з маніпулятивною риторикою
на основі лінгвістичних та екстралінгвістичних критеріїв**

Кваліфікаційна робота бакалавра
студентки 4 курсу
освітньої програми
*«Прикладна (комп'ютерна) лінгвістика
та англійська мова»*
спеціальності – 035.10 Філологія (прикладна
лінгвістика)
галузі знань – 03 гуманітарні науки
Марії КУЛДОШИНОЇ
Науковий керівник:
к.т.н., доц. Микола КОСТИКОВ

«Допущено до захисту»
Протокол засідання
кафедри української мови та прикладної лінгвістики
протокол № 15 від « 06 » 06 2024 року

завідувач кафедри _____ (підпис)
к.філол.н., доц. Сергій РІЗНИК

КИЇВ – 2024

АНОТАЦІЯ

У засобах повсякденного спілкування, таких як месенджери та соціальні мережі, користувачі щодня стикаються з маніпулятивними заявами, які в потоці інформації можуть сприйматися як істинні. В умовах російсько-української війни ця проблема є надзвичайно актуальною та гостро повстає питання підвищення рівня інформаційної безпеки громадян України.

Саме з цією метою було виконано поточне дослідження, спрямоване на створення програмного забезпечення, яке автоматизує процес розпізнавання маніпулятивної риторики в новинних Телеграм-каналах.

Об'єктом дослідження є лінгвістичні та екстралінгвістичні маркери маніпулятивної риторики, а предметом – комп'ютерна модель для їх ідентифікації.

Мета роботи: створення алгоритму для аналізу новинних текстів і виявлення в них ознак маніпуляцій. Для її досягнення аналізуються етичні стандарти журналістики та наявні методи розпізнавання дезінформації в текстах, а на основі отриманих даних визначаються маркери для автоматичної оцінки новинних публікацій.

Методологічну базу складають методи автоматичного аналізу текстів, комп'ютерного моделювання, морфологічного аналізу, екстралінгвістичного аналізу та статистичного аналізу. Таким чином, для визначення маніпулятивної риторики поєднуються лінгвістичні та технічні підходи.

За підсумками дослідження було розроблено програмне забезпечення та реалізовано його функціонал у Телеграм-боті. Практичне тестування підтвердило ефективність інструменту. Результати вказують на значний потенціал методики для підвищення інформаційної безпеки та підтримки медіаграмотності серед населення в умовах постійних інформаційних атак.

Таким чином, проведене дослідження робить важливий внесок у розвиток інструментарію для боротьби з ворожим інформаційним впливом у воєнний час, пропонуючи новітній підхід та доступний спосіб перевірки медіаконтенту.

Ключові слова: інформаційна безпека, маніпуляції в медіа, Телеграм-канали, морфологічний аналіз, статистичний аналіз, екстралінгвістичний аналіз, NLP.

ABSTRACT

In everyday communication mediums, such as messaging apps and social networks, users frequently encounter manipulative statements that may be perceived as truthful when within the information flow. In the context of the Russian-Ukrainian war, this issue is extremely relevant and raises the issue of enhancing the information security level of Ukrainian citizens.

This study was conducted with the aim of creating software that automates the process of recognizing manipulative rhetoric in news Telegram channels.

The objects of the research are the linguistic and extralinguistic markers of manipulative rhetoric, and the subject is a computer model for their identification. The goal of the study is to create an algorithm for analyzing news texts and detecting signs of manipulation. To achieve this, the ethical standards of journalism and existing methods for recognizing disinformation in texts are analyzed, and markers for the automatic evaluation of news publications are determined based on the collected data.

The methodological base includes methods of automatic text analysis, computer modeling, morphological analysis, extralinguistic analysis, and statistical analysis. Thus, linguistic and technical approaches are combined to identify manipulative rhetoric.

As a result of the study, software and its implementation in a Telegram bot were developed. Practical testing confirmed the effectiveness of the tool. The results indicate significant potential of the methodology for enhancing information security and supporting media literacy among the population in the face of constant information attacks.

Thus, the conducted research makes an important contribution to the development of tools for combating hostile information influence in wartime, offering a novel approach and an accessible way to verify media content.

Keywords: information security, media manipulation, Telegram channels, morphological analysis, statistical analysis, extralinguistic analysis, NLP.

ЗМІСТ

ВСТУП.....	6
РОЗДІЛ 1. ТЕОРЕТИЧНІ ЗАСАДИ ТА МАРКЕРИ МАНІПУЛЯТИВНОЇ РИТОРИКИ В ЗАСОБАХ МАСОВОЇ ІНФОРМАЦІЇ.....	11
1.1. Аналіз етичних стандартів сучасної журналістики.....	11
1.2. Методи маніпуляції та поширення дезінформації.....	12
1.3. Ключові лінгвістичні та екстралінгвістичні ознаки маніпуляцій і дезінформації в новинних текстах.....	16
РОЗДІЛ 2. СТВОРЕННЯ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ДЛЯ РОЗПІЗНАВАННЯ ТЕЛЕГРАМ-КАНАЛІВ З МАНІПУЛЯТИВНОЮ РИТОРИКОЮ.....	20
2.1. Наявні методи автоматичного розпізнавання маніпуляцій і дезінформації в новинах.....	20
2.2. Агрегація текстових матеріалів з анонімних і журналістських каналів.....	22
2.3. Створення алгоритму обчислення кількісних і статистичних показників маркерів маніпуляції.....	25
2.4. Аналіз статистичних даних та квантифікування критеріїв розпізнавання маніпулятивних текстів.....	30
2.5. Розробка програмного забезпечення для автоматичного розпізнавання новинних каналів з маніпулятивною риторикою та його реалізація в Телеграм-боті.....	37
2.6. Демонстрація роботи Телеграм-бота «zvirchannel_bot».....	41
ВИСНОВКИ.....	48
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	50
ЕЛЕКТРОННІ ДЖЕРЕЛА.....	54
СПИСОК ДЖЕРЕЛ ТЕКСТОВОГО МАТЕРІАЛУ.....	58
ДОДАТКИ.....	59
Додаток 1. Програмний код.....	59
Додаток 2. Текстові матеріали.....	60
Додаток 3. Результати статистичних обрахунків показників лінгвістичних і екстралінгвістичних ознак.....	61
Додаток 4. Вміст файлу «markers.csv».....	71

ВСТУП

Інформаційні операції є одним із невіддільних аспектів військових конфліктів. Вони використовуються для маніпуляції громадською думкою, зміни сприйняття та впливу на рішення як на національному, так і на міжнародному рівнях. Через велике розповсюдження соціальних медіа та цифрових комунікацій, інформаційні операції стали ще більш значущими та впливовими, і це успішно використовує Російська Федерація.

Дослідники-політтехнологи стверджують, що інформаційна війна Росії проти України розпочалась ще наприкінці ХХ століття, у часи розпаду СРСР [6], однак беззаперечно, що найяскравіший розквіт її припав на ХХІ століття. Багато розголосу отримували російські інформаційні операції здійснені під час Революції гідності, Антитерористичної операції на сході України й операції Об'єднаних Сил, але найбільше – ті, що проводяться з 2022 року.

Вони передували й надалі супроводжують широкомасштабне вторгнення Росії в Україну [24].

Так, зокрема, у 2022 році Росія активно фінансувала та використовувала канали в месенджері Телеграм, видаючи їх за українські та поширюючи в них вигідні для себе наративи [38]. Надалі використання месенджерів стало одним з ключових інструментів у російських інформаційних операціях. Ворог адаптувався, почав використовувати українську мову й активніше залучати ботоферми.

Однак сучасні російські інформаційні операції дедалі частіше реалізуються не лише країною-агресором і її агентами. Кропітке вивчення українського інформаційного простору та настроїв громадян дозволило Росії шляхом тонких маніпуляцій залучити до поширення шкідницьких наративів і самих українців, які нині підхоплюють, поширюють і роздмухують ворожі ідеї в соціальних мережах і месенджерах.

Сьогодні тисячі українців отримують інформацію з анонімних Телеграм-каналів, які, здавалося б, належать патріотам України, але насправді

активно поширюють інформацію, що часто є неповною, маніпулятивною чи геть неправдивою. І хоча мотивація авторів і бенефіціарів цих каналів може бути різною, від простого бажання заробити до справжньої роботи на ворога [27], кінцевий результат зазвичай подібний – такі новинні джерела сприяють розповсюдженню ворожих наративів, що підривають довіру до державних установ і лідерів країни та вносять розбрат у суспільство.

На цю проблему неодноразово звертали увагу як представники державних органів, Служби безпеки України, Генеральної розвідки України, так і окремі активісти та журналісти. Наразі в Україні діють численні проєкти й ініціативи, що досліджують роботу російських агентів і пропагандистів, а також навчають громадян базових принципів інформаційної гігієни [32; 29; 46]. Однак, попри всі спроби взяти ситуацію під контроль, перевага в інформаційній війні досі залишається на боці ворога.

Очевидно, що розв'язання проблеми систематичної дезінформації та маніпуляції в Інтернеті потребує всебічного підходу та залучення спеціалістів різного профілю, оскільки інформаційно-психологічні операції можуть проводитись різними каналами та поширюватися у різноманітних формах — через фотографії, аудіо, відео та тексти. У контексті цієї наукової роботи головна увага приділяється текстовій інформації, а саме – публікаціям україномовних новинних Телеграм-каналів.

Актуальність дослідження полягає в підвищенні рівня інформаційної безпеки українців під час російсько-української війни шляхом автоматизації процесу розпізнавання маніпуляційної риторики в новинних Телеграм-каналах.

Наукова проблема: розпізнавання маніпулятивних наративів є складним завданням через різноманітність методів, які використовуються для їх поширення; автоматизація цього процесу потребує їх кропіткого аналізу, визначення ключових маркерів маніпуляцій у текстах новин та квантифікації їх показників.

Мета дослідження: на основі лінгвістичних та екстралінгвістичних ознак Телеграм-каналів без і з маніпулятивною риторикою створити програмне забезпечення для її автоматичного розпізнавання.

Завдання дослідження:

- 1) провести аналіз етичних стандартів сучасної журналістики;
- 2) дослідити методи, що використовуються для маніпуляції та поширення дезінформації;
- 3) визначити ключові лінгвістичні та екстралінгвістичні ознаки, що вказують на маніпуляції та неправдивість інформації в новинних текстах;
- 4) дослідити наявні методи автоматичного розпізнавання маніпуляцій і дезінформації в новинах;
- 5) зібрати текстові матеріали з надійних і ненадійних Телеграм-каналів для порівняння показників;
- 6) розробити алгоритм та провести обчислення кількісних і статистичних показників маркерів маніпуляції в новинних текстах;
- 7) проаналізувати отримані статистичні дані та квантифікувати критерії розпізнавання маніпулятивних текстів;
- 8) розробити програмне забезпечення для автоматичного розпізнавання новинних каналів з маніпулятивною риторикою та реалізувати його в Телеграм-боті;
- 9) продемонструвати роботу Телеграм-бота.

Об'єктом дослідження є лінгвістичні та екстралінгвістичні маркери маніпулятивної риторики у текстах новинних Телеграм-каналів.

Предметом дослідження є комп'ютерна модель автоматичного розпізнавання маніпулятивної риторики у текстах новинних Телеграм-каналів.

Методики та методи дослідження:

- метод автоматичного аналізу – використання алгоритмів для автоматичного аналізу текстів з метою виявлення маніпуляцій;

- метод комп'ютерного моделювання – створення моделей для автоматичного розпізнавання маніпулятивних текстів;
- методика морфологічного аналізу – використання NLP-бібліотеки для аналізу морфологічних характеристик тексту;
- методика екстралінгвістичного аналізу – аналіз текстів на предмет використання гіперпосилань, емодзі та верхнього регістру;
- методика статистичного аналізу – застосування статистичних формул.

Новизна роботи: у межах дослідження було розроблено інноваційний для української мови підхід до виявлення маніпулятивної риторики в Телеграм-каналах, який поєднує аналіз як лінгвістичних, так і екстралінгвістичних ознак. Цей підхід дозволяє просто та ефективно розпізнавати недоброчесні джерела.

Практичне значення роботи полягає у можливості використання розробленого програмного забезпечення для автоматичного розпізнавання Телеграм-каналів з маніпулятивною риторикою, що сприятиме підвищенню рівня інформаційної безпеки користувачів та допоможе їм виявляти й уникати ненадійних джерел інформації. Результати дослідження можуть також бути застосовані для розробки інструментів моніторингу інформації в соціальних мережах.

Теоретичне значення: у ході роботи шляхом комбінації методик морфологічного та екстралінгвістичного аналізу розроблено новаторську методику розпізнавання ознак маніпуляційної риторики в Телеграм-каналах, а також виділено найпоказовіші з цих маркерів. Отримані результати сприяють розвитку методик лінгвістичного аналізу та автоматичного розпізнавання маніпуляцій у текстах.

Інформаційна база дослідження: мова програмування Python [39] та її бібліотеки: csv [43], pandas [56], spacy [59], re [58], emoji [47], collections [42], math

[51], matplotlib [52], telebot [60], asyncio [40], telethon [61], datetime [44], pytz [57], io [50].

Матеріалом дослідження послуговували дві генеральні сукупності текстових публікацій укладених на основі десяти новинних Телеграм-каналів (п'яти ненадійних анонімних і п'яти перевірених [28] журналістських) за період від 1 січня 2024 року по 30 квітня 2024 року, укладених у дві вибірки по 400000 слів кожна.

Апробація дослідження: доповідь «Можливості застосування комп'ютерної лінгвістики у військовій справі» на X Міжнародній науково-технічній Internet-конференції «Сучасні методи, інформаційне, програмне та технічне забезпечення систем керування організаційно-технічними та технологічними комплексами» [1].

РОЗДІЛ 1. ТЕОРЕТИЧНІ ЗАСАДИ ТА МАРКЕРИ МАНІПУЛЯТИВНОЇ РИТОРИКИ В ЗАСОБАХ МАСОВОЇ ІНФОРМАЦІЇ

1.1. Аналіз етичних стандартів сучасної журналістики

Серед численних визначень журналістики можна визначити наступне як найзагальніше: «це форма громадської та літературної діяльності по збиранню, обробці та поширенню інформації через канали масової комунікації, під якими розуміються передусім преса, радіомовлення, телебачення, інформаційні агентства, електронні канали зв'язку» [2].

Таким чином, до журналістської діяльності можна віднести й Телеграм-канали, які в сучасній медійній реальності починають грати все більш значущу роль. Месенджери, такі як Телеграм, уже стали головними каналами зв'язку та оперативного отримання інформації в Україні, тому журналісти все частіше використовують їх для дублювання новинної стрічки з сайтів видань. Це дозволяє швидко та ефективно досягати аудиторії, а також розширювати коло читачів.

Однак, використання Телеграм-каналів у журналістській діяльності також викликає певне занепокоєння у контексті верифікації інформації та відповідальності. Хоча офіційні канали журналістських видань зазвичай дотримуються стандартів журналістики, неінституціоналізовані канали окремих публічних осіб і анонімні канали, можуть не відповідати цим критеріям. Відсутність редакційного контролю в таких випадках може призвести до розповсюдження неперевіреної інформації, що є особливо ризикованим під час інформаційних війн або політичних криз.

Саме задля уникнення цих ризиків існують стандарти журналістської етики, які визначають правила та принципи збору, обробки та поширення інформації. В Україні журналісти керуються «Етичним кодексом українського журналіста»

затвердженим Національною спілкою журналістів України, який закликає до поваги до свободи слова, неупередженого висвітлення судових процесів, захисту приватного життя та наголошує на обов'язку журналіста надавати об'єктивну інформацію [30].

Інші стандарти також вимагають від журналістів надавати лише точну та перевірену інформацію [13], забороняють створювати рекламні та спонсорські матеріали [20], а для онлайн видань – рекомендують надавати покликання на джерела інформації та мати інформаційну сторінку «Про нас» для забезпечення прозорості [14].

Варто також зазначити, що під час військових конфліктів крім звичайних етичних стандартів на журналістів також накладаються вимоги військової цензури, які переважно стосуються утримання певної інформації або її аспектів від публікації [10]. Необережна робота з інформацією у воєнний час може поставити під загрозу життя людей або існування держави загалом.

Таким чином, справжня якісна та надійна журналістська робота полягає в дотриманні актуальних етичних стандартів і норм. У цифрову епоху, коли інформація поширюється миттєво, а неправдива інформація – ще швидше [26], важливість верифікації фактів та об'єктивності набуває особливого значення. Дотримання цих принципів не тільки сприяє збереженню довіри аудиторії, але й утверджує медіа як надійне та відповідальне.

1.2. Методи маніпуляції та поширення дезінформації

Після російського вторгнення в Україну 24 лютого 2022 року Телеграм став основним джерелом актуальної інформації для багатьох українців. Так, згідно з дослідженням USAID, у 2023 році 72% українців використовували саме цей месенджер для отримання новин [3]. З розширенням загального кола читачів і ростом попиту на інформаційний контент збільшувалась і кількість новинних

Телеграм-каналів, значна частина з яких була та залишається анонімними. Вони й стали майданчиками для проведення російських інформаційних операцій і поширення дезінформації.

Неінституціоналізовані анонімні канали можна прирівняти до електронної версії «жовтої преси» – сенсаційної журналістики [17]. Таким джерелам не властива об'єктивність і неупередженість. Керуючись власними інтересами та вигодою замість журналістських стандартів, їх власники не гребують популізмом, провокаційними заголовками та маніпуляціями задля створення сенсаційного контенту.

Оскільки це явище не нове, воно має формальне визначення навіть у юриспруденції. Так одним з трактувань маніпуляцій у засобах масової інформації (ЗМІ) є «процес створення образу чи аргументації на користь певних інтересів», іншим – «використання соціальних мереж для поширення оманливої або неточної інформації» [15].

На практиці, існують численні методи, що використовуються для маніпулювання інформацією, кожен з яких має свої специфічні цілі та механізми впливу на громадську думку. Розглянемо деякі з цих методів, спираючись на статтю Поліщук Ю.Я., Гнатюк С.О. і Сейлової Н.А. «Мас медіа як канал маніпулятивного впливу на суспільство» [4].

- Методи «отруйного» і «цукрового сендвіча» – специфічне розташування інформації, що має на меті приховати невігідні дані. Результат досягається шляхом подачі позитивного повідомлення між двома негативними у випадку «отруйного сендвіча» і навпаки – шляхом маскуванню негативної інформації між блоками хорошої у «цукровому». Таким чином знижується важливість і вплив проміжного повідомлення.
- Метод «буденної розповіді» – спокійна й рутинна подача негативної інформації, що має на меті призвичаєння людини до негативу та зменшення емоційної реакції на нього.

- Метод «міфотворення» – розповсюдження неправдивих наративів, які приймаються населенням без критичного осмислення завдяки регулярному повторенню та емоційному навантаженню.
- Метод «анонімного авторитету» – посилення на неназвані авторитети задля підтвердження аргументів; слугує для введення аудиторії в оману й завоювання її довіри.
- Метод «емоційного резонансу» або «сенсаціалізації» – специфічна подача інформації з підсвічуванням і драматизацією її окремих аспектів, що викликає сильний емоційний відгук у читача [12]. Така включеність забезпечує уважне та беззаперечне сприйняття аудиторією упереджених і однобоких даних, а також подальше його активне поширення [11].
- Метод «створення інформаційної реальності» – відображення фактів у вигідній джерелу інтерпретації або шляхом вибіркового їх освітлення; спрямований на формування конкретного сприйняття подій аудиторією, привернення уваги до одних і замовчування інших.
- Метод «забовтування» – підвищення уваги до певної теми та частотне її обговорення, часто з перебільшенням деталей, що зрештою призводить до перенасичення аудиторії інформацією, втоми та спаду інтересу або формування негативного сприйняття цієї теми.
- Метод «інформаційної блокади» – замовчування інформації, що не вписується в наративну лінію. Таким чином дані, які можуть спричинити шкоду, не досягають громадськості.
- Метод «коментарів» – представлення інформації разом з інтерпретацією джерела, у якій акцентуються окремі аспекти або подається вигідний джерелу аналіз. Метод слугує для формування громадської думки шляхом її спрямування.

- Метод «ефекту ореолу» – паралелізація інформації зі згадками відомих осіб або подій з метою створення сприятливого або несприятливого іміджу, заснованого на асоціаціях.
- Метод «ефекту первинності» – оперативна комунікація свіжих новин задля формування первісних вражень, які визначатимуть подальшу інтерпретацію прокомунікованих даних. Метод заснований на психологічній тенденції людей вірити інформації, яку вони почули першою.
- Метод «класифікаторів» – використання узагальнень і кліше задля створення швидких асоціацій або стереотипів, які спрощують інформаційний потік для розуміння та створюють підґрунтя для ефективного навіювання.
- Метод «констатації факту» – представлення сценаріїв або заяв як доконаних фактів з метою формування певних настроїв у суспільстві без реальних на то підстав.
- Метод «помилкової аналогії» – порівняння двох різних об'єктів або ситуацій, які мають мінімальну або жодної спільності, що вводить аудиторію в оману та підштовхує до формування висновків на основі некоректних порівнянь.
- Метод «зворотного зв'язку» – залучення читачів або глядачів до дискусії з метою направлення ходу думок аудиторії у вигідне річище. Метод часто реалізується за участі підставних осіб або ботів, які задають загальний ідейний тон полеміки [23]. Під тиском такої монобільшості, люди, що спочатку були в опозиції, схиляються до сумнівів або кардинально змінюють власну думку, що неодноразово підтвердив у своїх дослідженнях конформізму Соломон Аш [7; 8; 9].

Окремо в контексті онлайн-медіа варто також згадати метод «клікбейту» – заголовків, що презентують новину як сенсацію, упускають ключові деталі, містять особисті звертання або ставлять питання, які мають заінтригувати читача [16]. Первісно клікбейтами (від англ. «наживка на клацання») називались заголовки-покликання, які спонукають читачів натиснути на них, щоб прочитати

повну статтю. Однак у контексті месенджерів клікбейтами варто вважати заголовки текстів, які користувач бачить на екрані діалогів.

Іншими способами маніпуляції онлайн є використання емодзі, що є модифікацією методу «емоційного резонансу», та розміщення замовних повідомлень у новинній стрічці, зокрема без відповідного маркування [33].

Хоча медіа завжди мали значний вплив на колективну свідомість, розвиток технологій значно посилив їхню здатність маніпулювати громадською думкою, яка часто сприймає інформацію некритично та легко піддається зовнішнім впливам. Оскільки більшість читачів не вдається до глибокого теоретичного аналізу, їхні погляди легко формуються під впливом зовнішніх чинників та інформаційного тиску. А високий рівень маніпуляції дозволяє недоброчесним медіа нав'язувати громадськості ідеї, підсилюючи цей вплив.

1.3. Ключові лінгвістичні та екстралінгвістичні ознаки маніпуляцій і дезінформації в новинних текстах

Необхідним етапом для виконання практичної частини дослідження є аналіз лінгвістичних та екстралінгвістичних ознак, які потенційно можуть вказувати на маніпуляції та неправдивість новинних текстів. Їх ідентифікація дозволяє визначати викривлення інформації та забезпечує інструментарій для оцінки об'єктивності новинного контенту.

До групи лінгвістичних ознак можна віднести будь-які специфічні мовні особливості, які використовуються для маніпулювання сприйняттям читача. У цьому дослідженні як мовні маркери розглядаються морфологічні категорії.

Екстралінгвістичні ознаки, своєю чергою, стосуються контекстуальних і візуальних аспектів інформації. У межах цієї групи в дослідженні розглядаються не мовні елементи текстів, які можуть бути визначені під час обробки тексту програмним кодом.

Для початку розглянемо, які маркери маніпуляції в текстах визначають дослідники. Зокрема науковиці Констанкевич І.М., Костусяк Н.М., Шульська Н.М., Станіслав О.В., Єлова Т.А. та Кауза І.Б. у статті про маніпуляції в медіа [21] серед інших вказують на наступні ознаки:

- численне повторення фраз і слів, які привертають увагу читача («терміново»), підсилюють емоційну реакцію («шок») або викликають довіру («надійне джерело») і можуть служити спробою закріпити неправдиву інформацію у свідомості аудиторії;
- використання плеоназму («смертельно вбитий труп») та гіперболізації для перебільшення фактів і провокування емоційної реакції;
- персоналізація та введення елемента особистої залученості за допомогою фраз «ви знаєте мене, я не брешу», «військовий мені повідомив» задля створення враження правдивості інформації;
- створення повідомлень на основі думок і оціночних суджень, без реальних фактів, а також їх формулювання в емоційному або істеричному тоні;
- недотримання елементарних правил орфографії, порушення лінгвістичних стилістичних норм, використання русизмів або інших лексичних невідповідностей.

Ще однією маніпулятивною методикою є використання простих пропозицій, наприклад простих речень, що складаються з головних членів речення та доповнення [5]. Це забезпечує легке сприйняття та засвоєння текстів читачами.

Спираючись на перелічені вище ознаки маніпулятивної риторики можна виділити декілька лінгвістичних та екстралінгвістичних маркерів для аналізу:

- 1) списки слів, які використовуються для привернення уваги аудиторії, сенсаціоналізації новини та провокування емоційної реакції;
- 2) ознаки гіперболізації, зокрема списки слів, що можуть підсилювати значення інших лексем, та прикметники найвищого ступеня порівняння;
- 3) засоби персоналізації контенту, а саме займенники першої та другої особи;

- 4) некоректне написання лексичних одиниць
- 5) кількість слів у реченнях і використання пунктуації та сполучників як ознаки складності синтаксичних пропозицій.

Також, спираючись на результати роботи Капусти Й., Гаєка П., Мунка М. і Бенка Л. [19], які досліджували лінгвістичні ознаки неправдивих новин, можемо виділити декілька морфологічних ознак, які також могли б послугувати ідентифікаторами маніпулятивної риторики. Дослідження показало, що в новинах від недоброчесних джерел спостерігається підвищена частота використання іншомовних слів, прикметників та іменників. Водночас у правдивих новинах було зафіксовано вищу частоту вживання питальних і відносних займенників, а також детермінативів (артиклів, присвійних та вказівних займенників), прийменників і дієслів.

Варто зазначити, що неправдиві (фейкові) новини, маркери яких аналізувались у вищезгаданому дослідженні, і маніпулятивна риторика не є тотожними поняттями. Маніпуляції в ЗМІ є широким означенням застосування психологічного впливу для зміни поведінки або сприйняття даних аудиторією. В цьому контексті можна виділити ще вузчі категорії: дезінформацію – цілеспрямоване поширення хибної чи оманливої інформації; мізінформацію – ненавмисне розповсюдження шкідливих повідомлень; і малінформацію – навмисне поширення правдивих фактів з метою завдання шкоди. Неправдиві новини, які з'являються в результаті дез- і мізінформування, є підкатегоріями медійних маніпуляцій, тому їх ознаки можуть виявитись корисними для ідентифікації маніпулятивних практик.

Окрім того, спираючись на результати дослідження Інституту Масової Інформації [32] на тему маніпуляцій в Телеграм-каналах, до екстралінгвістичних маркерів можна також зарахувати використання емодзі та покликань. Перші, як згадувалось у [підрозділі 1.2.](#), є модифікацією методу «емоційного резонансу», а другі – активно використовуються в рекламних цілях.

Зрештою, ще одним маркером маніпулятивності джерела, який нині прослідковується в неінституціоналізованих Телеграм-каналах, є використання верхнього регістру. Найчастіше у месенджерах ним виділяють заголовки новин – частину тексту, яка з’явиться на екрані діалогів Телеграм (мал 1.1.). Таким чином, канал виокремлюється серед інших чатів, провокуючи в користувача відчуття важливості та терміновості повідомлення. Цей метод фактично є адаптацією клікбейт-заголовків до середовища месенджерів.



Малюнок 1.1. Скриншот клікбейтного заголовка на екрані діалогів Телеграм.

Отже, у цьому підрозділі було розглянуто різні лінгвістичні та екстралінгвістичні ознаки, які можуть вказувати на маніпуляцію та неправдивість інформації в новинах. Лінгвістичні ознаки охоплюють морфологічні категорії та інші мовні особливості, які сприяють маніпуляції сприйняттям читача. Серед них варто звернути увагу на деякі з частин мови (іменники, прикметники, займенники, прийменники та дієслова) та їх окремі мовні парадигми (перша та друга особа займенників, найвищий ступінь порівняння прикметників). Екстралінгвістичні ознаки представлені контекстуальними та візуальними аспектами текстів. Серед них як ключові виділено використання емодзі, покликань і верхнього регістру.

Ці знахідки можуть бути використані для ідентифікації потенційно маніпулятивних чи неправдивих новин.

РОЗДІЛ 2. СТВОРЕННЯ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ДЛЯ РОЗПІЗНАВАННЯ ТЕЛЕГРАМ-КАНАЛІВ З МАНІПУЛЯТИВНОЮ РИТОРИКОЮ

2.1. Найвні методи автоматичного розпізнавання маніпуляцій і дезінформації в новинах

Ідея автоматичного виявлення неправдивої інформації не є новою, але досі залишається надзвичайно актуальною в контексті сучасного інформаційного простору. З огляду на значний вплив маніпулятивних новин на громадську думку та політичні процеси, розробка та впровадження ефективних технологій для їхнього автоматичного розпізнавання стає ключовим завданням для науковців, технологів та політичних аналітиків.

Хоча окремі спеціалісти досі намагаються реалізувати цю задачу вручну, шляхом маркування обраних лінгвістичних аспектів і їхнього подальшого обрахунку за допомогою програм [22], більшість дослідників нині схиляється до використання спеціалізованих інструментів. Розглянемо деякі з таких робіт і методи, які в них застосовуються.

У дослідженні Капусти Й., Хаска П., Мунка М. і Бенка Л. від 2020 року було проаналізовано відмінності між неправдивими та правдивими новинами за допомогою морфологічного аналізу [19]. Дослідники зібрали датасет новин, за допомогою інструменту TreeTagger [64] виконали морфологічний аналіз для визначення мовних особливостей текстів, виділили ключові лінгвістичні ознаки, такі як частота слів та структури речень, і порівняли статистичні показники для ідентифікації істотних відмінностей. На основі цих ознак були натреновані моделі машинного навчання для диференціації між реальними та фейковими новинами.

Дослідження проведене на базі Університету Північної Дакоти використало частоти частин мови та N-грам (послідовностей з двох або більше символів чи

слів) для виявлення дезінформації [25]. За допомогою бібліотек NLTK [54] дослідники проаналізували розподіл різних частин мови (зокрема іменників, дієслів та прикметників) у новинах і, використавши результати аналізу отриманих частот, покращили точність роботи нейронної мережі, що розпізнає неправдиві новини, до понад 80%. Для класифікації статей вони також застосували N-грами до п'яти символів; в результаті була отримана стабільна точність моделі близько 60%. Таким чином, ефективність обидвох лінгвістичних інструментів, частот частин мови та N-грам, у визначенні дезінформації була доведена.

Інше дослідження за участі Капусти Й., Дрліка М. і Мунка М. застосувало метод N-грам до тегів частин мови для виявлення дезінформації, зокрема неправдивих новин, пов'язаних з COVID-19 [18]. Основні методи, що використовувались у дослідженні, включають морфологічний аналіз частин мови (дієслів і різних типів іменників) за допомогою TreeTagger і застосування N-грам, згенерованих на основі отриманих тегів частин мови. Дослідження показало, що результати поєднання морфологічних методів з методом N-грам порівнянні з результатами аналізу з використанням TF-IDF (статистичного показника, що вимірює важливість слова в контексті документа на тлі корпусу) за точністю класифікації.

Розвиток машинного навчання та штучного інтелекту відкрив нові можливості для аналізу великих обсягів даних, що дозволяє більш ефективно ідентифікувати та контролювати розповсюдження маніпулятивних повідомлень. Тут варто згадати й про рішення, які вже були запроваджені та зараз функціонують як доступні користувачам сервіси, що дозволяють перевірити англomовну інформацію на правдивість.

1. Одним із них є «FactCheck.org» [48], незалежний інтернет-ресурс, який спеціалізується на верифікації політичних заяв та новин. Ресурс перевіряє факти та публікує результати на своєму сайті, але не завжди встигає реагувати на події в реальному часі.

2. Іншим прикладом є «Snopes» [62], один з найавторитетніших ресурсів перевірки фактів. Він опрацьовує широкий спектр інформаційних тверджень, від буденних порад до політичних заяв, і об'єктивно оцінює тексти на надійність і правдивість. Цей сервіс також не завжди встигає оперативно опрацьовувати новини.
3. Третім подібним сервісом є «NewsGuard» [63] – розширенням для браузера, що використовує штучний інтелект для оцінки надійності новинних джерел, оцінюючи їх на основі точності, прозорості та інших критеріїв. Розширення інформує користувачів про надійність сайтів і попереджає, якщо джерело містить дезінформацію.

Таким чином, проаналізувавши деякі з наявних сервісів та методів, можемо зазначити, що на цей день вже було реалізовано чимало результативних дослідницьких проєктів та програмних рішень, спрямованих на автоматичне розпізнавання дезінформації в новинах. Проте ця глобальна проблема досі залишається не до кінця вирішеною та потребує подальшого вивчення та застосування нових підходів для її розв'язання.

Це практичне дослідження є продовженням роботи попередників. За ключові критерії аналізу в ньому приймаються морфологічні ознаки лексем, оскільки ефективність їх використання як маркерів дезінформації вже неодноразово була підтверджена іншими науковцями [22; 19; 18]. Це дослідження спрямоване на подальший розвиток та удосконалення методів виявлення маніпуляцій, а також на створення надійного й ефективного інструмента для боротьби з дезінформацією в медіапросторі.

2.2. Агрегація текстових матеріалів з анонімних і журналістських каналів

Першим кроком для імплементації порівняльного аналізу стала агрегація двох генеральних сукупностей текстів Телеграм-каналів – надійних і ненадійних.

Надійними вважались джерела журналістських видань, які в результаті двоетапного глибинного моніторингу онлайн-медіа Інститутом Масової Інформації були визнані прозорими та відповідальними й були внесені до «білого списку» [28]. Зокрема, з одинадцяти запропонованих каналів до аналізу було обрано наступні: «Бабель», «Еспресо», «Громадське», «Суспільне» й «Українська правда».

Ненадійними натомість вважались неінституціоналізовані анонімні джерела, у публікаціях яких неодноразово була помічена маніпулятивна риторика. З них, на противагу «білому списку», у межах дослідження було сформовано «сірий». До нього було внесено чотири канали з мережі «Труха», ґрунтовне дослідження про яку у вересні 2023 року було опубліковано на сайті NGL.media [36], та канал «Харків 1654», який за своєю сутністю є аналогічним до джерел згаданої вище мережі. Таким чином до «сірого списку» дослідження потрапили: «Адвокат Мілер», «Батальйон Монако», «Доброго вечора, ми з України», «Харків 1654» та «Труха Україна».

Оскільки система автоматичного розпізнавання маніпулятивної риторики в межах цього дослідження опрацьовує не окремі повідомлення, а загальний тон комунікації Телеграм-каналу, наступним важливим кроком стало визначення періоду, публікації за який будуть завантажені для подальшого порівняльного аналізу. Так, наприклад, файли з повідомленнями за цілий рік виявились занадто великими для ефективного опрацювання, оскільки об'єм даних деяких із них перевищував 20 мегабайтів. А результати аналізу публікацій за один місяць не відображали повноцінної лінгвістичної та екстралінгвістичної картини. Зрештою, дослідним шляхом було визначено, що оптимальним є період у чотири місяці.

Задля забезпечення контекстуальної стабільності експерименту період було визначено в календарних датах – від 1 січня 2024 року по 30 квітня 2024 року. Зважаючи на політичні події обраних місяців та загальний тривожний настрій у

суспільстві, очікувалось, що недоброчесні Телеграм-канали відображатимуть і підсилюватимуть емоційну картину серед громадськості.

Процес агрегації текстів за обраний період було реалізовано в локальному текстовому редакторі Visual Studio Code [53] мовою програмування Python. Повний код знаходиться в файлі «message_extraction.py» у теці, яку можна переглянути за покликанням у Додатку 1, пункті 2.

Програма складається з двох функцій: `extract_messages`, яка зосереджена на асинхронній взаємодії з зовнішнім API (від англ. «інтерфейс програмування застосунків») для збору даних, і `get_messages` – її синхронної обгортки. Перша функція відповідає за з'єднання з сервером, отримання повідомлень від каналу, та обробку цих повідомлень – усі ці дії оптимально виконувати асинхронно, оскільки очікування на відповідь сервера може займати довгий час. Тим часом друга функція дозволяє управляти асинхронними завданнями, як-от їх запуск і зупинка, що особливо важливо при використанні зовнішніх API. Зрештою, використання комбінації функцій не лише оптимізувало роботу програми, але й дозволило повторно використати її в роботі Телеграм-бота.

У коді «message_extraction.py» використовуються кілька допоміжних бібліотек:

- `asuncio` – стандартна бібліотека, використана для успішної роботи асинхронного коду за допомогою суміжних програм;
- `telethon` – стороння бібліотека для взаємодії з API Телеграму, яка дозволяє автоматизувати взаємодію з месенджером;
- `pandas` – високорівнева бібліотека для аналізу даних, яка використовується для створення об'єкта `DataFrame` [55] з завантажених даних та їх збереження у форматі `CSV`;
- `re` – стандартна бібліотека для роботи з регулярними виразами, застосована для визначення назви каналу з `URL`;

- `datetime` – стандартна бібліотека Python для роботи з датами й часом, яка використовується для визначення періоду, за яких збираються повідомлення;
- `pytz` – бібліотека для роботи з часовими зонами, за допомогою якої час повідомлень приводиться до київського.

У результаті роботи програми було створено 10 CSV файлів, однойменних до назв Телеграм-каналів, кожен з яких містив усі публікації певного джерела за період від 1 січня 2024 року по 30 квітня 2024 року. Повідомлення в CSV-файлах було укладено у дві колонки: «timestamp» з датами та часом повідомлень та «text» з текстами публікацій. Файли можна переглянути в теці за покликанням у Додатку 2. Саме на основі двох генеральних сукупностей, укладених з текстових матеріалів цих файлів, надалі проводилось обчислення кількісних і статистичних показників ознак маніпуляції.

2.3. Створення алгоритму обчислення кількісних і статистичних показників маркерів маніпуляції

На відміну від модулів, що використовуються для взаємодії з Телеграм, реалізованих у локальному редакторі Visual Studio Code, код алгоритму обчислення кількісних і статистичних показників маркерів маніпуляції розроблявся у хмарному середовищі редактора Google Colab [49]. Такий вибір зумовлений зручністю інтерфейсу сервісу, особливо вигідного для графічного представлення статистичних даних і графіків.

Код було організовано у дванадцять модулів: «Необхідні бібліотеки та файли для аналізу», «Укладання вибірок та підвибірок без (raw) та з (preprocessed) передобробкою», «Функції лінгвістичної та екстралінгвістичної обробки», «Функції визначення статистичних показників», «Підрахунки для покликань», «Підрахунки для емодзі», «Підрахунки для слів у верхньому регістрі», «Підрахунки для частин мов», «Підрахунки для першої, другої та третьої осіб

займенників», «Підрахунки для порівняльних ступенів прикметників», «Функції квантифікування маркерів маніпулятивної риторики» та «Запис числових меж маркерів анонімних та надійних каналів до CSV». Алгоритм для обчислення кількісних і статистичних показників маркерів маніпуляції в новинних текстах було реалізовано в перших чотирьох модулях програми. Повний код можна переглянути в теці за покликанням у Додатку 1, пункт 1.

У межах першого модуля було завантажено та під'єднано допоміжні бібліотеки, зокрема вже згадані вище `pandas` і `re`, а також:

- `csv` – стандартна бібліотека для імпортування та експортування даних у форматі CSV;
- `emoji` – бібліотека, яка дозволяє аналізувати емодзі в тексті;
- `collections`: стандартна бібліотека, модуль `Counter` якої допомагає підрахувати кількості кожного унікального елемента в ітерабельному об'єкті;
- `math`: стандартна бібліотека, що містить функції для математичних обчислень вищого рівня;
- `matplotlib`: бібліотека, модуль `pyplot` якої використовується для створення графіків і діаграм;
- `spacy` – передова NLP-бібліотека, що містить набір інструментів для аналізу мовних структур, виявлення іменованих сутностей, аналізу залежностей тощо, зокрема для української мови (модель «`uk_core_news_lg`»).

Також у першому модулі було визначено списки «`grey_news_files`» та «`white_news_files`», що містять назви CSV-файлів з публікаціями з анонімних та надійних журналістських каналів відповідно.

У модулі «Укладання вибірок та підвибірок без (`raw`) та з (`preprocessed`) передобробкою» було згруповано функції для конвертації завантажених з Телеграм матеріалів у словники Python. Первинна обробка реалізується у «`form_dict_raw`». Функція обробляє списки CSV-файлів, розсортовує матеріали кожного з каналів на чотири окремі групи за місяцем публікації та додатково розбиває отримані дані за

місяць на чотири менші групи. Унікальні ідентифікатори для відібраних повідомлень формуються шляхом комбінування перших трьох літер назви каналу, місяця публікації (01, 02, 03 або 04) та індексу групи від 1 до 4 (наприклад, «esp02_4» для повідомлень медіа «Еспресо» за останню чверть лютого). Таким чином, створюються словники двох генеральних сукупностей, з публікаціями анонімних джерел та надійних журналістських каналів, ключами в яких є ідентифікатори, а значеннями ключів – одна чверть необроблених повідомлень опублікованих за місяць, довжина яких коливається від 7000 до 15000 слів.

Далі вводиться функція «remove_trade_phrases», яка прибирає рядки самореклами, що повторюються наприкінці кожного повідомлення часто з закликком прислати новину або підписатися. Наявність самореклами властива дев'яти з десяти аналізованих Телеграм-джерел. У ній зокрема вживаються власні назви, використовується верхній регістр і емодзі, що може значно повпливати на показники подальшого лінгвістичного та екстралінгвістичного аналізу.

Окремо задається функція попередньої обробки тексту «preprocess_text», у межах якої застосовується «remove_trade_phrases», прибирається пунктуація, цифри, зайві пробіли, покликання й емодзі, великі літери замінюються на маленькі, а текст лематизується за допомогою бібліотеки spacy. Така передобробка дозволяє стандартизувати текст, забезпечує більшу однорідність і точність аналізу даних

На функцію «preprocess_text» посилається наступна – «form_dict_preprocessed», у межах якої довжини текстів повідомлень для кожного ключа спочатку скорочуються до 7000 слів, а потім запускається їх передобробка. Таким чином зменшується навантаження на систему, яке спричиняє застосування інструментів NLP-бібліотек.

Останньою функцією модуля є «form_subsamples», яка обробляє словники, формуючи остаточні текстові підвибірки в 5000 слів кожна.

Таким чином, після застосування всіх функцій у межах модуля з двох генеральних сукупностей текстів – маніпулятивних з анонімних каналів і не маніпулятивних з журналістських – було сформовано дві вибірки по 400000 слів кожна. Ці вибірки своєю чергою були розбиті на 80 підвбірок по 5000 слів, з дотриманням принципу рівномірного розподілу по генеральній сукупності, шляхом вилучення даних з чвертей кожного з чотирьох місяців. Такі вибірки формуються окремо для оброблених і не оброблених текстів і використовуються залежно від аналізованої ознаки.

У третьому модулі «Функції лінгвістичної та екстралінгвістичної обробки» визначаються функції підрахунку ознак, що потенційно можуть стати маркерами маніпулятивних риторик, зокрема:

- «calculate_links» – підрахунок кількості внутрішніх (на публікації та канали в Телеграм) і зовнішніх покликань за допомогою регулярних виразів;
- «calculate_emojis» – підрахунок кількості емодзі за допомогою бібліотеки emoji;
- «remove_prop_nouns» – допоміжна функція вилучення власних назв за допомогою бібліотеки spacy;
- «calculate_capitalization» – підрахунок кількості слів у верхньому регістрі за допомогою методу isupper [65]; перед обробкою до тексту застосовується функція «remove_prop_nouns» для вилучення абревіатур, з метою підрахунку саме емотивних прикладів капіталізації;
- «calculate_pos» – підрахунок лексем кожної з частин мов за допомогою бібліотеки spacy;
- «calculate_pers_pronouns» – підрахунок займенників першої, другої та третьої особи за допомогою бібліотеки spacy;
- «calculate_comparative_superlative» – підрахунок прикметників вищого та найвищого ступеня порівняння за допомогою бібліотеки spacy.

Останній модуль алгоритму визначає математичні функції. Так «calculations_to_df» укладає частотний словник і представляє дані будь-яких лінгвістичних або екстралінгвістичних підрахунків у вигляді таблиці, а «calculate_stats» реалізує всі необхідні статистичні підрахунки та виводить результати в термінал. Зокрема, на основі частотного словника певної ознаки будується варіативний ряд (x_i), обраховується кількість варіант (n_i) та їх добуток ($x_i \cdot n_i$). Окремим стовпчиком прораховуються квадратичні відхилення. У підсумку функція повертає показники середньої частоти, середнього квадратичного відхилення та довірчих інтервалів σ , 2σ та 3σ .

Створений і готовий для використання алгоритм було надалі застосовано для обчислення кількісних та статистичних показників для маркерів маніпуляції у вибірках новинних текстів. Обрахунки було організовано в шість модулів, по одному для покликань, емодзі, слів у верхньому регістрі, частин мов, займенників першої, другої та третьої осіб, а також вищого та найвищого ступенів порівняння прикметників. Повний код обрахунків можна переглянути в теці за покликанням у Додатку 1, пункт 1.

Алгоритм проведення обчислень залишався незмінним у всіх модулях. Першою до аналізованої ознаки застосовувалась функція «calculations_to_df», яка виводила в термінал таблицю абсолютних частот певної ознаки або її форм у кожній з підвбірок окремо маніпулятивних і окремо надійних текстів. Другою викликала функція «calculate_stats», за допомогою якою будувалась стовпчикова діаграма для варіаційного ряду ознаки та обчислювались ключові статистичні показники. У випадку коли функція лінгвістичної або екстралінгвістичної обробки аналізувала одразу декілька форм однієї ознаки, наприклад рахувала частоти внутрішніх і зовнішніх покликань або кожної з частин мови, «calculate_stats» застосовувалась до потенційних маркерів окремо.

Таким чином, алгоритм для обчислення кількісних і статистичних показників було реалізовано в чотирьох модулях. Було завантажено та під'єднано необхідні

для роботи бібліотеки та матеріали, визначено функції для укладання вибірок, для лінгвістичного та екстралінгвістичного аналізу, а також для проведення статистичних підрахунків. Надалі цей алгоритм було застосовано для обчислення кількісних та статистичних показників потенційних маркерів маніпуляції у підвибірках маніпулятивних і надійних текстів. Результати обчислення є основою для подальшої роботи, зокрема для квантифікування критеріїв розпізнавання маніпулятивної риторики.

2.4. Аналіз статистичних даних та квантифікування критеріїв розпізнавання маніпулятивних текстів

У результаті обчислення кількісних і статистичних показників потенційних маркерів маніпуляції в новинних текстах кожна з ознак або її форм отримала стислу статистичну характеристику та була представлена графічно шляхом візуалізації варіативного ряду. Отримані дані потребували детального аналізу, метою якого було виділити релевантні маркери маніпулятивної риторики й прийняти довірчу ймовірність, з якою надалі могли б бути ідентифіковані недоброчесні джерела. Квантифікування критеріїв розпізнавання реалізовувалось на основі результатів аналізу.

Вибір довірчої ймовірності вимагав попереднього порівняння смуг коливань кожного показника в маніпулятивних і надійних текстах. У процесі аналізу було виявлено, що деякі інтервали не мали статистично значущих розходжень і були майже ідентичні, а отже не могли б послугувати інформативними критеріями для відрізнення маніпулятивних текстів від не маніпулятивних. Такими, зокрема, були довірчі інтервали, отримані для іменників (табл. 2.1.), дієслів (табл. 2.2.) та сполучників сурядності й підрядності (Додаток 3), тому всі три частини мови було визнано неінформативними та вилучено зі списку маркерів для подальшого аналізу.

Таблиця 2.1. Середні частоти та довірчі інтервали для іменників у маніпулятивних і надійних джерелах.

	Середня частота	Довірчий інтервал σ	Довірчий інтервал 2σ	Довірчий інтервал 3σ
Маніпулятивні джерела	1688.15	1622.04 - 1754.26	1555.93 - 1820.37	1489.82 - 1886.48
Надійні джерела	1711	1653.86 - 1768.14	1596.72 - 1825.28	1539.58 - 1882.42

Таблиця 2.2. Середні частоти та довірчі інтервали для дієслів у маніпулятивних і надійних джерелах.

	Середня частота	Довірчий інтервал σ	Довірчий інтервал 2σ	Довірчий інтервал 3σ
Маніпулятивні джерела	629.99	581.65 - 678.33	533.31 - 726.67	484.97 - 775.01
Надійні джерела	635.04	604.6 - 665.48	574.16 - 695.92	543.72 - 726.36

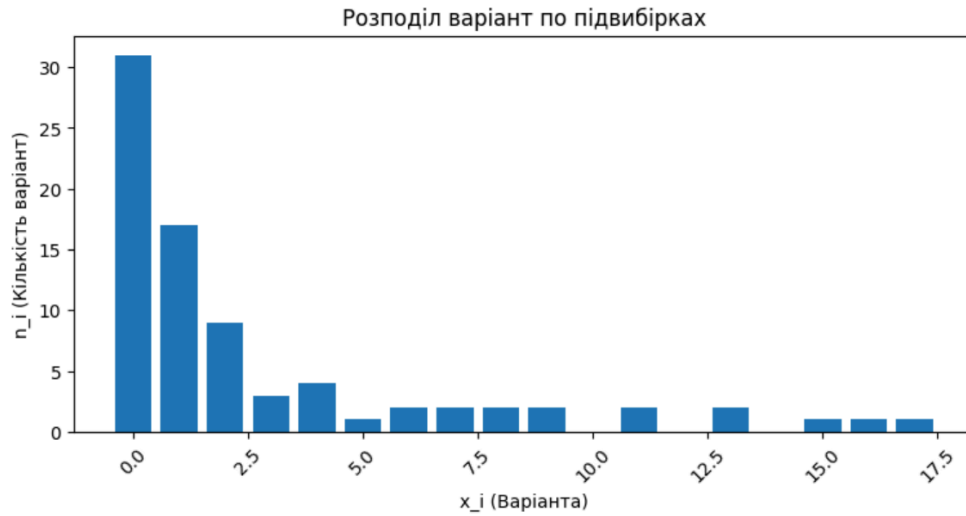
У випадку інших показників, таких як використання покликань або емодзі, інтервали значно відрізнялися, що свідчило про інформативність цих ознак та дозволило включити їх до списку маркерів для розпізнавання маніпулятивних текстів. У табл. 2.3., як приклад інформативних результатів, наведено середні частоти та межі довірчих інтервалів, обчислених для внутрішніх і зовнішніх покликань, а мал. 2.1. і мал. 2.2. ілюструють варіативні ряди для зовнішніх покликань, представлені у вигляді стовпчикових діаграм.

Таблиця 2.3. Середні частоти та довірчі інтервали для внутрішніх і зовнішніх покликань у маніпулятивних і надійних джерелах.

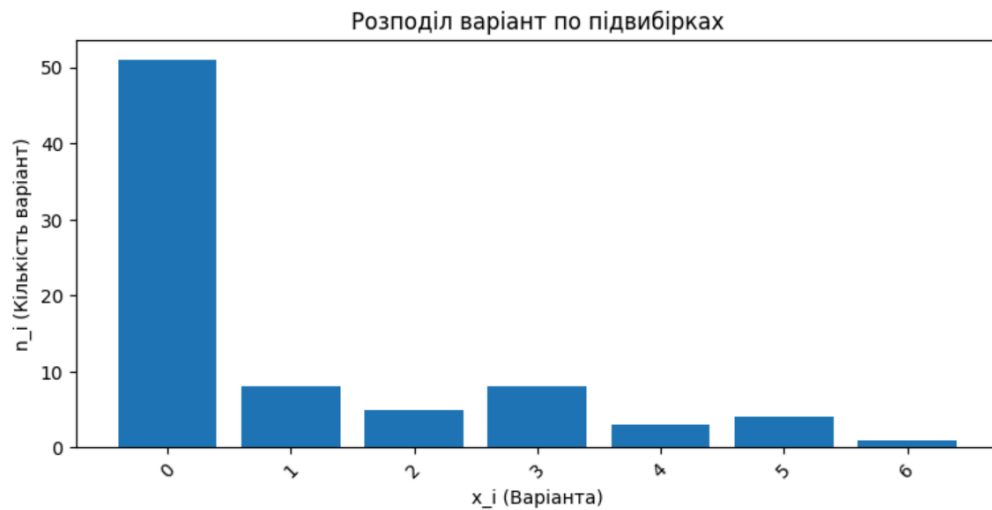
	Середня частота	Довірчий інтервал σ	Довірчий інтервал 2σ	Довірчий інтервал 3σ
Внутрішні покликання				
Маніпулятивні джерела	5.81	-0.87 - 12.49	-7.55 - 19.17	-14.23 - 25.85
Надійні джерела	0.01	-0.1 - 0.12	-0.21 - 0.23	-0.32 - 0.34

Продовження табл. 2.3. Середні частоти та довірчі інтервали для внутрішніх і зовнішніх покликань у маніпулятивних і надійних джерелах.

Зовнішні покликання				
Маніпулятивні джерела	2.76	-1.38 - 6.9	-5.52 - 11.04	-9.66 - 15.18
Надійні джерела	1	-0.6 - 2.6	-2.2 - 4.2	-3.8 - 5.8



Малюнок 2.1. Варіативний ряд для зовнішніх покликань у маніпулятивних джерелах.



Малюнок 2.2. Варіативний ряд для зовнішніх покликань у надійних джерелах.

Частотне використання покликань, як і очікувалось, спостерігалось саме у вибірці матеріалів з маніпулятивною риторикою, зокрема зазвичай – у рекламних цілях (мал. 2.3.).

Варто зауважити, що під час експорту повідомлень з Телеграм-каналів не зберігались такі елементи форматування тексту, як виділення жирним шрифтом, курсивом, підкреслення, вбудовані покликання та інші методи. Отже, показники обчислювались саме для гіперпосилань прописаних у тексті.



Малюнок 2.3. Скриншот рекламного повідомлення на анонімному каналі «Батальйон Монако»

Високий рівень інформативності також мали наступні ознаки: використання емодзі та верхнього регістру, вживання власних назв, прикметників і їх найвищого ступеня порівняння, прислівників, числівників, займенників, особливо першої та

другої особи, та часток. Довірчі інтервали отримані для прикметників вищого ступеня порівняння та займенників третьої особи в обох вибірках були майже ідентичними, тому ці ознаки також виявились неінформативними. З усім тим, було вирішено не вилучати їх зі списку маркерів для збереження повноти парадигм на етапі презентації результатів користувачу.

Усі таблиці з середніми частотами та межами довірчих інтервалів для проаналізованих ознак, а також їх варіативні ряди, представлені у вигляді стовпчикових діаграм, можна переглянути в Додатку 3.

Таким чином, після аналізу статистичних даних для розпізнавання Телеграм-каналів з маніпулятивною риторикою були обрані наступні лінгвістичні та екстралінгвістичні маркери: внутрішні та зовнішні покликання, емодзі, використання верхнього регістру, власні назви, прикметники, вищий і найвищий ступінь порівняння прикметників, прислівники, числівники, займенники, особові займенники першої, другої та третьої осіб, а також частки.

Довірчі інтервали σ для більшості з перелічених ознак у вибірках маніпулятивних і не маніпулятивних текстів не перетинались, що свідчить про неістотне розходження середніх частот. Однак, зважаючи на те, що тексти обох вибірок написані в одному стилі, а їх розрізнення проводиться за слабо вираженою ознакою, дослідження продовжуватиметься на основі наявних даних із застосуванням додаткових функцій для розмежування інтервалів.

На основі аналізу отриманих результатів було прийнято рішення обрати інтервали 2σ для квантифікування маркерів текстів з маніпулятивною риторикою, що мало б забезпечити точність визначення з довірчою ймовірністю 95,5%. Були також внесені зміни до функції «calculate_stats» алгоритму, описаного в [підрозділі 2.3.](#), щоб оновлена функція повертала списком граничні значення довірчого інтервалу 2σ .

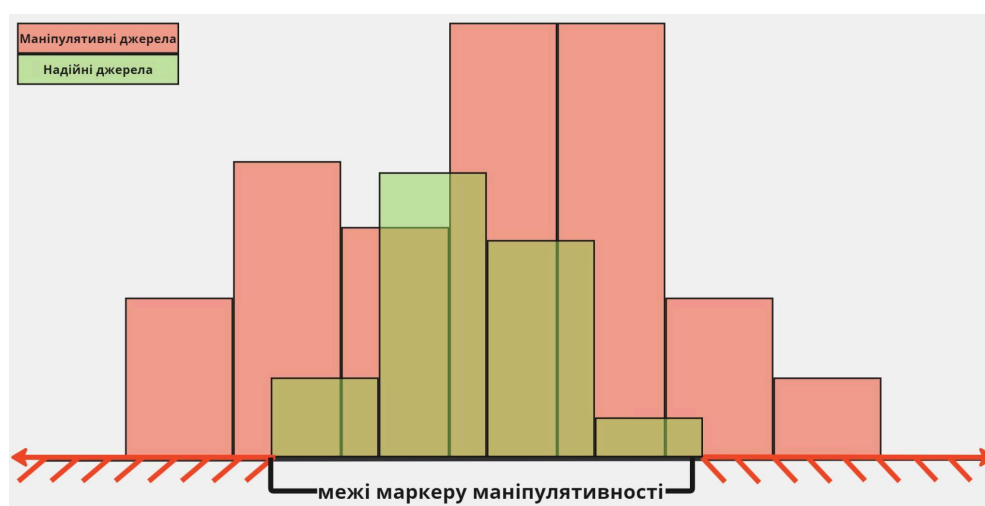
Наступним кроком стало квантифікування маркерів, за допомогою яких можна було б визначити, чи є аналізований текст маніпулятивним. Ідея полягала в тому, щоб на основі довірчого інтервалу 2σ обрахувати середні частоти, які могли б бути властиві лише текстам ненадійних джерел.

Ця задача була реалізована в межах модуля «Функції квантифікування маркерів маніпулятивної риторики», який містить методи «`intervals_to_markers`» і «`compile_markers`». Перша функція, шляхом порівняння меж довірчих інтервалів 2σ , обрахованих для певної ознаки в маніпулятивних і не маніпулятивних вибірках, визначає граничні значення, що можуть послугувати маркерами ненадійності текстів.

Друга ж функція, «`compile_markers`», збирає отримані квантифіковані маркери в список словників з ключами «`parameter`» (ознака), «`grey_min`» (ліва межа маркеру маніпулятивної риторики) і «`grey_max`» (права межа того ж маркеру). Після аналізу такий список може бути записаний у CSV-файл.

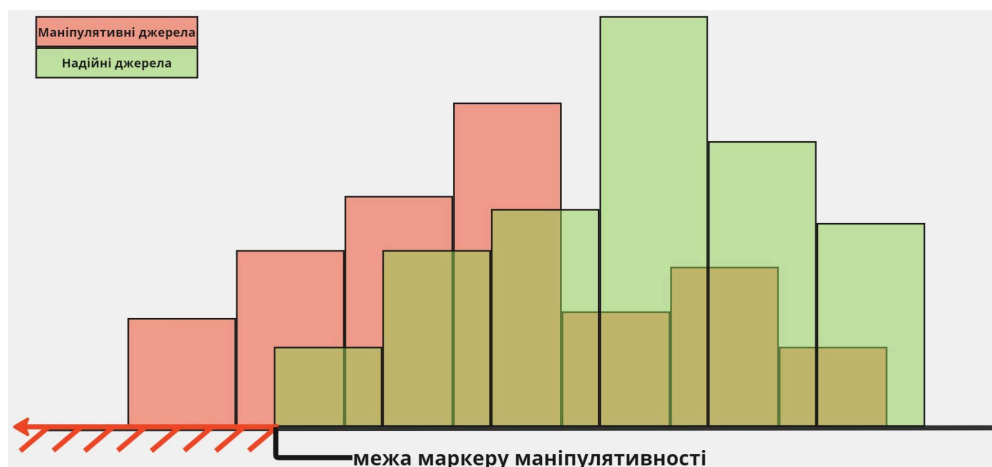
Порівняльний аналіз функції «`intervals_to_markers`» заснований на простому алгоритмі з трьох кроків:

- Якщо довірчий інтервал надійних текстів знаходиться в межах довірчого інтервалу маніпулятивних текстів, граничні значення других приймаються за маркери маніпулятивності (мал. 2.4.).



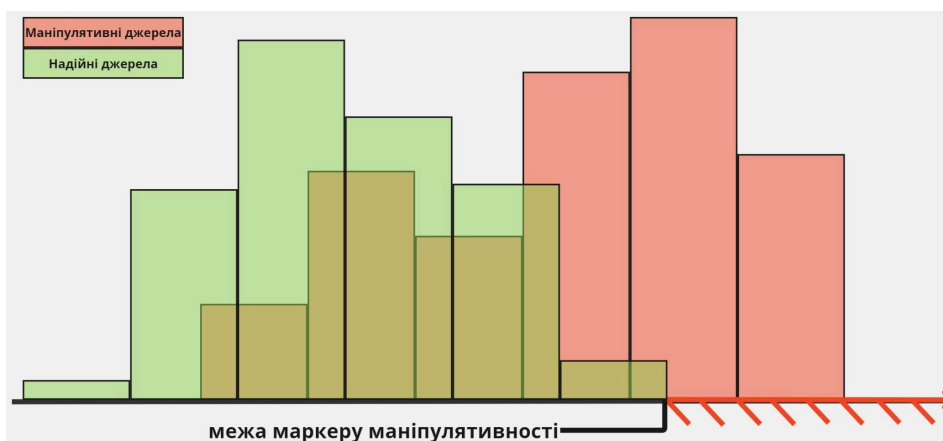
Малюнок 2.4. Ілюстрація двосторонніх меж маркеру маніпулятивності.

- Якщо нижня межа граничного інтервалу маніпулятивних текстів менша за нижню межу надійних і верхня межа маніпулятивних також менша за верхню межу надійних, тоді приймається лише ліва межа маркерів маніпулятивності – нижнє граничне значення надійних текстів (мал. 2.5.).



Малюнок 2.5. Ілюстрація лівої межі маркеру маніпулятивності.

- Якщо нижня межа граничного інтервалу маніпулятивних текстів більша за нижню межу надійних і верхня межа маніпулятивних також більша за верхню межу надійних, тоді приймається лише права межа маркерів маніпулятивності – верхнє граничне значення надійних текстів (мал. 2.6.).



Малюнок 2.6. Ілюстрація правої межі маркеру маніпулятивності.

Обидві функції модуля були застосовані до всіх відібраних лінгвістичних і екстралінгвістичних маркерів маніпулятивної риторики, і на основі отриманих результатів було укладено список «markers_list», що містив ліві та праві межі всіх

квантифікованих маркерів. Цей список за допомогою бібліотеки csv було збережено як CSV-файл у модулі «Запис числових меж маркерів анонімних та надійних каналів до CSV». З вмістом укладеного файлу, що надалі був використаний для розпізнавання Телеграм-каналів з маніпулятивною риторикою, можна ознайомитись у Додатку 4.

Таким чином, аналітична стадія проекту була завершена. Були створені необхідні статистичні та NLP-функції, а також деякі допоміжні; були виконані всі необхідні обчислення та визначені маркери, що можуть послугувати для ідентифікації Телеграм-каналів з маніпулятивною риторикою, а їх межі було квантифіковано й укладено в словник «markers.csv». Зібраних даних було достатньо для застосування їх на практиці.

2.5. Розробка програмного забезпечення для автоматичного розпізнавання новинних каналів з маніпулятивною риторикою та його реалізація в Телеграм-боті


Для створення, запуску та підтримки роботи Телеграм-бота, який автоматично розпізнає Телеграм-канали з маніпулятивною риторикою, було знову використано редактор Visual Studio Code. Вибір його, замість Google Colab, пояснюється безпековими міркуваннями. Робота з API Телеграму вимагає використання персональних ідентифікаторів, які надійніше зберігати та запускати в локальному редакторі, аніж у хмарному середовищі.

Програмне забезпечення для автоматичного розпізнавання новинних каналів з маніпулятивною риторикою було укладено в п'ять модулів: «message_extraction», «functions_linguistics», «functions_processing», «functions_markers_check» і «zvirchannel_bot». Модуль «message_extraction» був детально описаний в [підрозділі 2.2.](#) та лише повторно використовується у функціоналі Телеграм-бота.

Він завантажує публікації з новинного каналу, покликання на який надає користувач, за період від 1 січня 2024 року по 30 квітня 2024 року.

Модуль «functions_linguistics» містить функції лінгвістичної та екстралінгвістичної обробки, про які йшлося в [підрозділі 2.3.](#), зокрема: «remove_prop_nouns», «calculate_links», «calculate_emojis», «calculate_capitalization», «calculate_pos», «calculate_pers_pronouns», «calculate_comparative_superlative», а також функцію попередньої обробки тексту «preprocess_text», яка також використовує NLP-бібліотеку spacy. Таке групування дозволяє зменшити навантаження на систему, шляхом виконання NLP-обробки в межах одного модуля.

Решта функцій, що використовуються для укладання вибірок та підвибірок і визначення статистичних показників, зібрані в модулі «functions_processing». Серед них зокрема: «form_dict_raw», «form_dict_preprocessed», «form_subsamples», «calculate_stats». Усі чотири функції також були детально описані в [підрозділі 2.3.](#) У межах функціонала Телеграм-бота вони так само відповідають за попередню обробку генеральних сукупностей текстів і створення двох вибірок, маніпулятивних і не маніпулятивних текстів, по 80 підвибірок у 5000 слів кожна, а також – проведення статистичних обрахунків.

Четвертий модуль, «functions_markers_check», відповідає за перевірку текстів на наявність ознак маніпулятивної риторики за допомогою маркерів з файлу «markers.csv», про який йшла мова наприкінці [підрозділу 2.4.](#) На початку модуля, за допомогою бібліотеки pandas, квантифіковані значення маркерів завантажуються у словник «grey_markers». Наступною задається функція «apply_markers», яка позначає квантифіковану ознаку аналізованого тексту символом «» як маніпулятивну, якщо її середня частота у вибірці знаходиться за межами інтервалу, визначеного маркерами зі словника «grey_markers», або нижче лівої чи вище правої межі у випадках, коли визначена лише одна з них.



Друга функція модуля, «metrics_print», є допоміжною. Вона приймає на вхід назву аналізованої ознаки, словник текстової вибірки, функцію лінгвістичної або екстралінгвістичної обробки, яка має бути застосована до ознаки, коментар українською, який побачить користувач, і назву маркеру ознаки в словнику «grey_markers». На вихід функція повертає назву ознаки в коментарі, її середню частоту в текстах вибірки та, за допомогою «apply_markers», позначає або не позначає її як маніпулятивну, залежно від результатів порівняння середньої частоти з квантифікованими маркерами.

Остання функція, «check_channel_against_metrics», викликає решту функцій чотирьох модулів, формує оброблені та необроблені словники текстових вибірок, та визначає повідомлення-відповідь користувачу з результатами опрацювання матеріалів каналу. У межах цієї функції багаторазово викликається «metrics_print» для аналізу показників кожної з ознак. Тут також використовується бібліотека io, яка дозволяє перенаправляти вихідний текст функції напряму в Телеграм-бот, без створення тимчасових файлів у пам'яті системи.

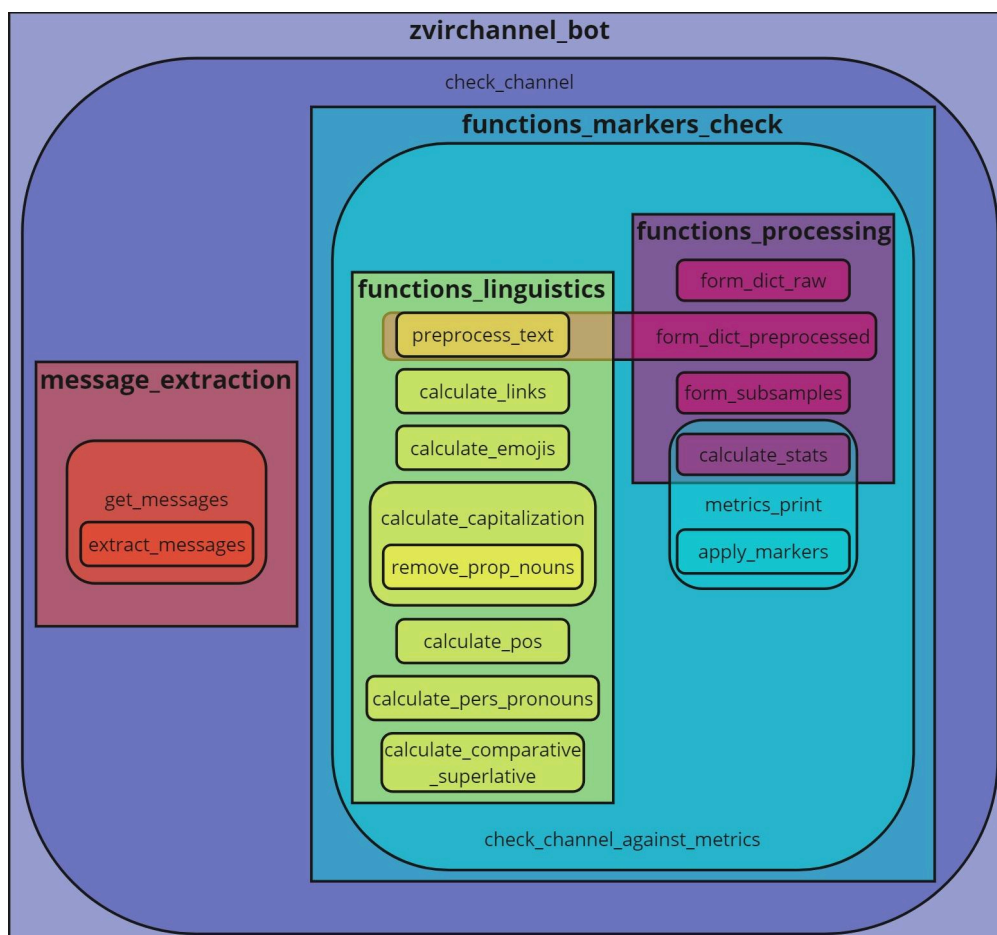
Останнім і головним є модуль «zvirchannel_bot». Саме він запускає, координує та підтримує роботу Телеграм-бота. Цей модуль використовує бібліотеку telebot для під'єднання до месенджера та взаємодії з ботом. Алгоритм роботи бота дуже простий, з трьома лінійними гілками. Після запуску програми командою «/start» користувач отримує вітальне повідомлення та запрошення натиснути на одну з трьох кнопок, які скеровують його на відповідну гілку.

1. Гілка «Перевірити канал 🐾» відправляє користувачу запит на покликання на канал та попереджає, що обробка матеріалів триватиме 5-10 хвилин.

Процес запускається функцією «check_channel», яка з модуля «message_extraction» викликає «get_messages» для завантаження текстових матеріалів, а з модуля «functions_markers_check» – функцію «check_channel_against_metrics», що обробляє текст та повертає результат перевірки.

2. Гілка «Критерії перевірки » розлого інформує користувача, які ознаки аналізуються програмою та чому саме вони були обрані.
3. Гілка «Про недоброчесні канали » надає інформацію про базові принципи інформаційної гігієни під час війни, розповідає про небезпеки анонімних Телеграм-каналів і надає покликання на джерела з білого списку перевірених медіа.

Таким чином було сформовано всі п'ять модулів програмного забезпечення Телеграм-бота, структуру яких представлено на схемі нижче (мал. 2.7.). Квадрати та прямокутники відображають модулі, назви яких подано жирним шрифтом, а заокруглені форми – функції цих модулів. Вони можуть вміщати в собі інші фігури-функції, якщо в коді посилаються на них.



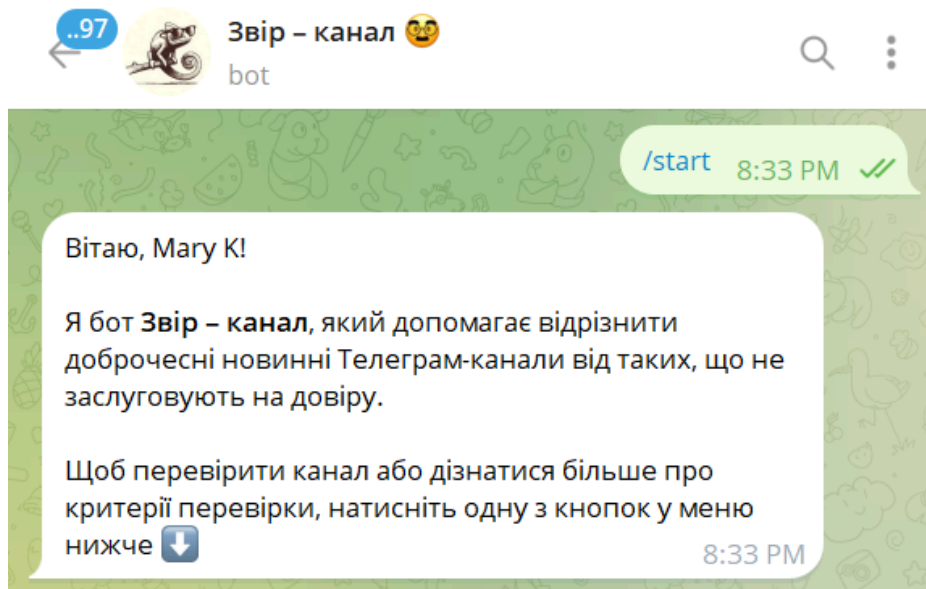
Малюнок 2.7. Схема взаємодії модулів і функцій Телеграм-бота «zvirchannel_bot».

Перед запуском програми, було використано Телеграм-бот «BotFather» [41], який дозволяє створювати власних чат-ботів. Назва була визначена як «Звір – канал», а ім'я користувача – як «zvirchannel_bot». Після отримання цих даних «BotFather» надав унікальний токен, який було використано в модулі «zvirchannel_bot» для під'єднання до Телеграм-бота.

Варто зауважити, що Телеграм-бот «Звір – канал» активний лише тоді, коли програмний код запущений на комп'ютері, оскільки без робочого програмного середовища, яке підтримує його функціонування, бот не може взаємодіяти з серверами Телеграм або обробляти запити від користувачів. Щоб забезпечити постійну роботу ботів, незалежно від стану локального обладнання, використовуються зовнішні хостинги, наприклад, хмарні сервери або спеціалізовані сервіси для ботів. Однак це потребує додаткових фінансових витрат на підтримку хостингу. Саме тому на цьому етапі реалізації проєкту було вирішено користуватись локальним хостингом на персональному комп'ютері, що, однак, не забезпечує безперервного функціонування бота.

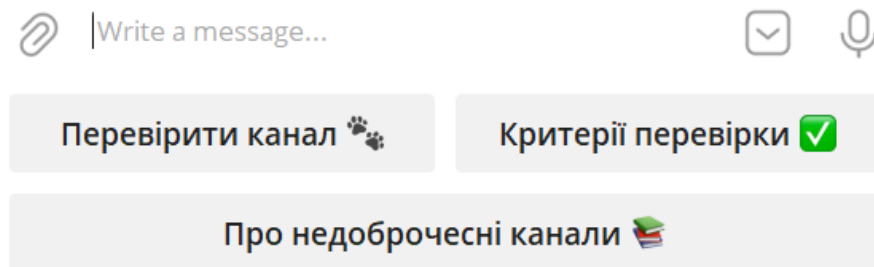
2.6. Демонстрація роботи Телеграм-бота «zvirchannel_bot»

Запуск Телеграм-бота. Щоб запустити взаємодію з ботом, користувач відправляє в чат команду «/start». У відповідь на неї він отримує вітальне повідомлення, в якому бот звертається до користувача за іменем, вказаним у Телеграм, і запрошує натиснути одну з кнопок для продовження взаємодії (мал. 2.8.).



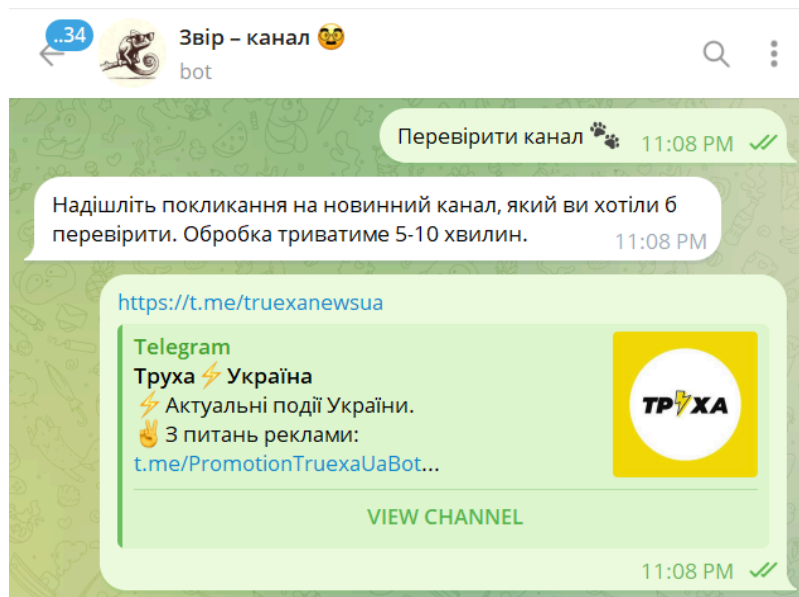
Малюнок 2.8. Скриншот початку взаємодії з Телеграм-ботом.

Кнопки розташовані під полем вводу повідомлення (мал. 2.9.). Залежно від вибору користувача, він може продовжити взаємодію за гілкою «Перевірити канал 🐾», «Критерії перевірки ✓» або «Про недобродесні канали 📖».




Малюнок 2.9. Скриншот кнопок бота.


Гілка «Перевірити канал 🐾». Натиснувши на кнопку, користувач отримує запит на надсилання покликання на Телеграм-канал, який він хотів би перевірити, а також попередження про тривалість обробки матеріалу. У відповідь користувач надсилає гіперпосилання скопійоване з опису каналу (мал. 2.10.).



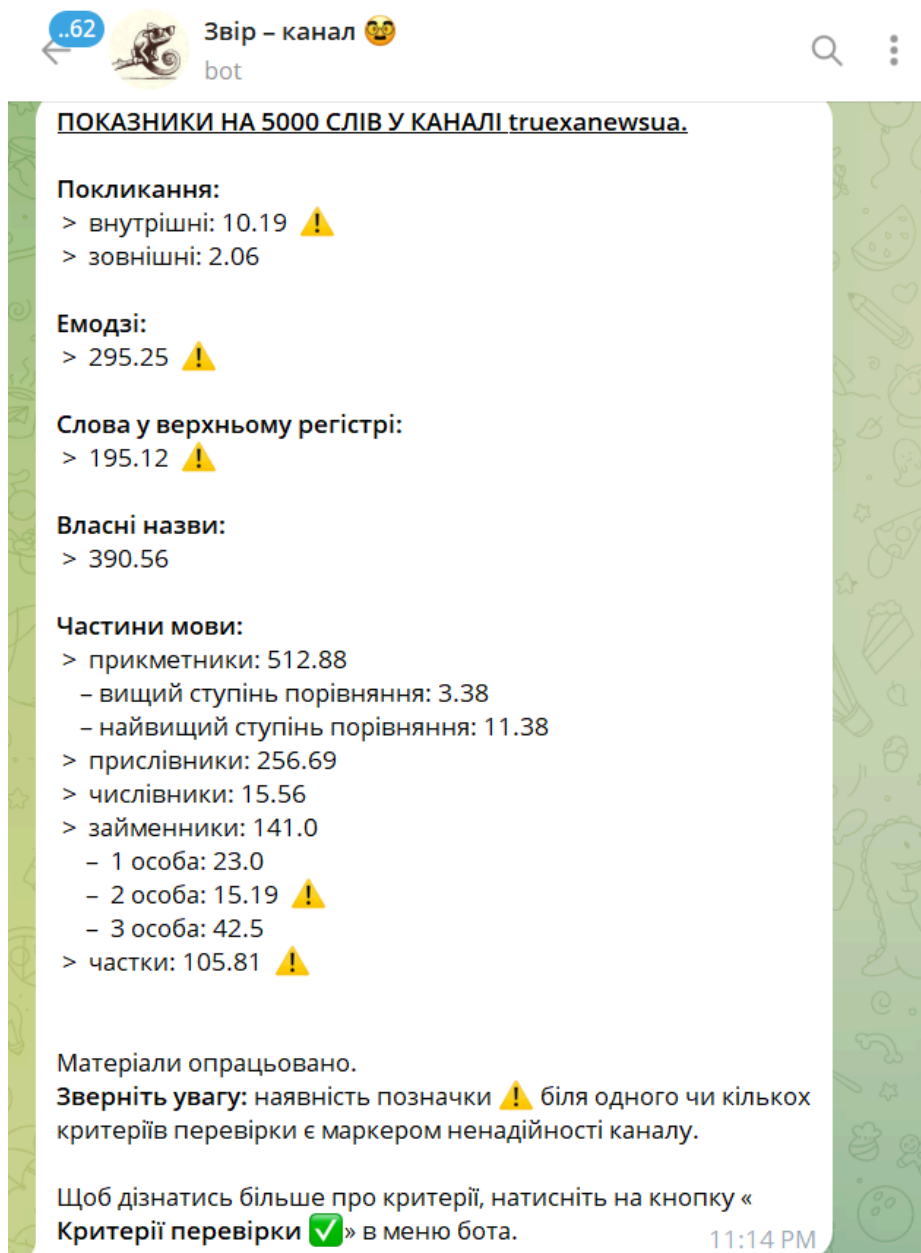
Малюнок 2.10. Скриншот запиту на перевірку Телеграм-каналу.

Залежно від об'єму матеріалу за 5-10 хвилин користувач отримує повідомлення від бота, в якому для кожної з проаналізованих ознак подається її середня частота вживання на 5000 слів. Показники, що були визначені як маркери маніпулятивної риторики позначаються символом «⚠️». Наприкінці повідомлення бот повідомляє, що наявність позначки свідчить про ненадійність каналу, та запрошує до подальшої взаємодії (мал. 2.11.).

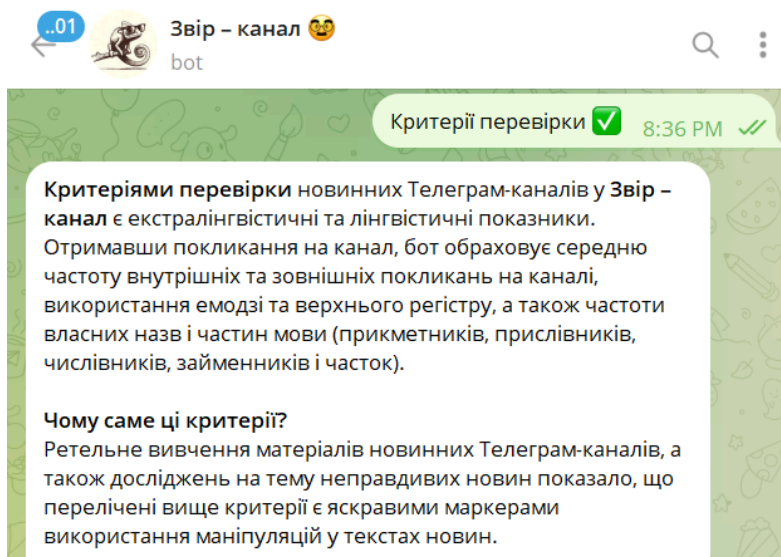
Гілка «Критерії перевірки» . Натиснувши на кнопку, користувач отримує розлоге повідомлення, в якому пояснюється принцип його роботи (мал. 2.12.) та вибір критеріїв аналізу (мал. 2.13.). Наприкінці повідомлення бот запрошує користувача до подальшої взаємодії.

Гілка «Про недоброчесні канали» . Натиснувши на кнопку, користувач отримує розлоге повідомлення з інформацією про виклики спричинені російською інформаційною війною, зокрема ті, що повстали після російського вторгнення у 2022 році, зокрема увага звертається на проблему анонімних каналів, які поширюють дезінформацію та експлуатують страхи людей, часто допомагаючи ворогу (мал. 2.14.). Як альтернатива анонімним каналам бот пропонує користувачу

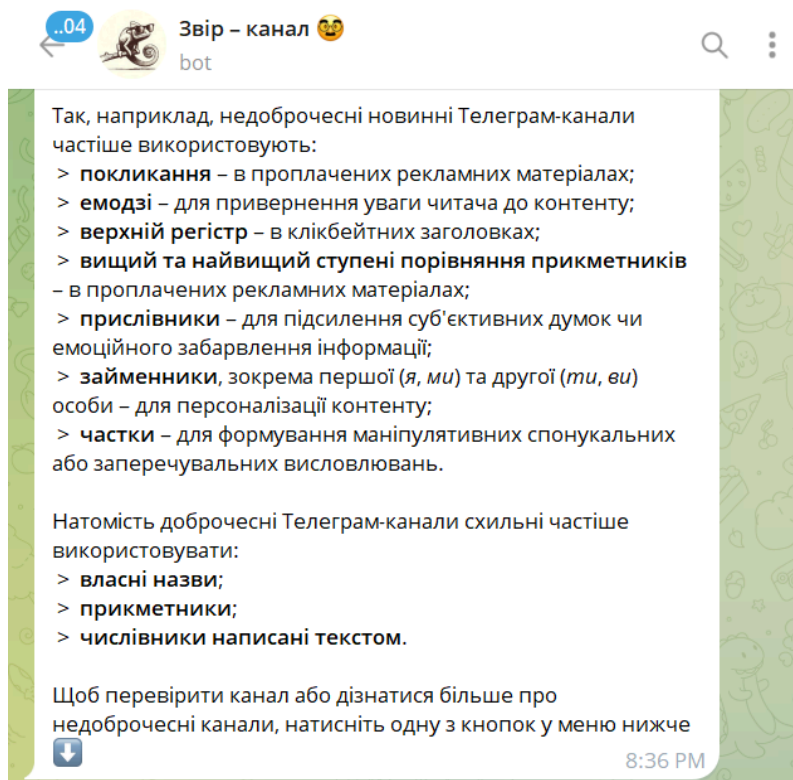
білий список перевірених медіа з покликаннями на кожне (мал. 2.15). Наприкінці повідомлення бот запрошує користувача до подальшої взаємодії.



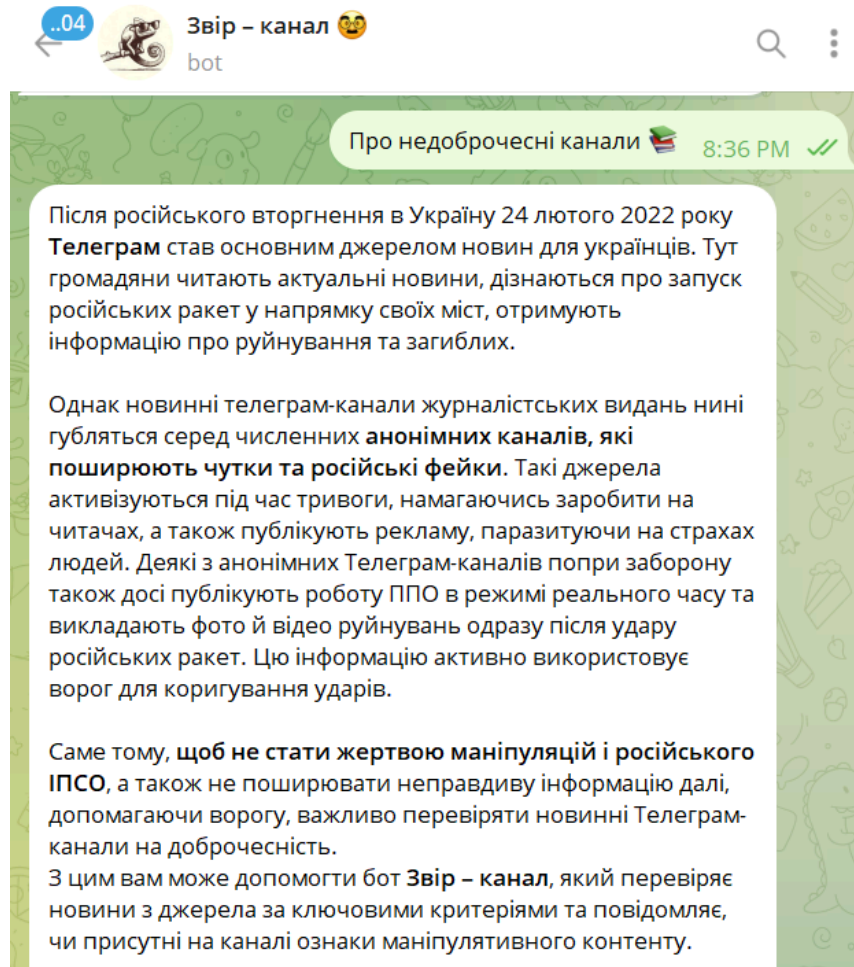
Малюнок 2.11. Скриншот результатів перевірки Телеграм-каналу.



Малюнок 2.12. Скриншот повідомлення-реакції на запит «Критерії перевірки ✓»: інформація про принцип роботи бота.



Малюнок 2.13. Скриншот повідомлення-реакції на запит «Критерії перевірки ✓»: маркери маніпулятивності.



Малюнок 2.14. Скриншот повідомлення-реакції на запит «Про недоброчесні канали 📖»:
інформаційні операції та роль анонімних каналів у них.

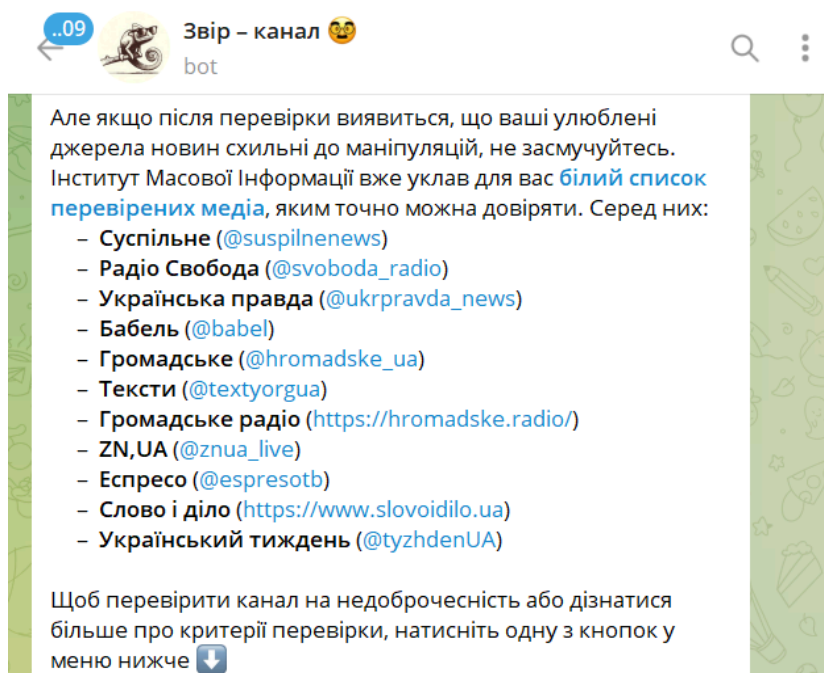
Перевірка якості. Телеграм-бот було протестовано на трьох анонімних каналах і трьох – з «білого списку» з метою визначення точності роботи програмного забезпечення. Показники двох надійних джерел, «Радіо Свобода» [34] і «Український тиждень» [37], були визнані надійними, а в «ZN.UA» [66] було виявлено надмірне використання емодзі.

Натомість перевірка анонімних каналів виявила численні маркери маніпуляцій в текстах публікацій, зокрема такими були:

- кількість зовнішніх покликань, власних назв, прикметників і займенників першої та другої особи на каналі «DeepStateUA» [45];
- кількість внутрішніх покликань, власних назв, прикметників, прислівників і часток на каналі «Радник» [35];

- кількість внутрішніх покликань і власних назв на каналі «Дріт» [31], який, однак, не вважається маніпулятивним.

Таким чином, ефективність роботи Телеграм-бота була доведена на практиці.



Малюнок 2.15. Скриншот повідомлення-реакції на запит «Про недоброчесні канали»: білий список перевірених медіа.

ВИСНОВКИ

Здійснене дослідження мало на меті створити програмне забезпечення для автоматичного визначення Телеграм-каналів з маніпулятивною риторикою. У результаті роботи було розроблено інноваційний для української мови підхід, який поєднує аналіз лінгвістичних і екстралінгвістичних ознак, обрахунки їх статистичних показників і визначення на основі зібраних даних маркерів ознак маніпуляцій у текстах новин. Функціонал програми було реалізовано в Телеграм-боті. Цей підхід дозволив просто та ефективно розпізнавати недоброчесні джерела в месенджері, що потенційно може підвищити рівень інформаційної безпеки українців, зокрема в часи російсько-української війни.

У межах роботи було проведено попереднє теоретичне дослідження етичних стандартів сучасної журналістики та методів, що використовуються для маніпуляції та поширення дезінформації у неправдивих новинах і жовтій пресі. На основі цієї інформації було визначено ключові лінгвістичні та екстралінгвістичні ознаки, що могли б вказувати на маніпуляції та неправдивість публікацій.

У практичній частині роботи було проаналізовано наявні методи автоматичного розпізнавання маніпуляцій і дезінформації в новинних текстах. Ефективність морфологічного методу була підтверджена багатьма дослідженнями, і саме його було застосовано в цій роботі. У численних наукових працях для розпізнавання неправдивих новин було застосовано метод машинного навчання, а в деяких він комбінувався з іншими методами. Отже, можемо припустити, що використання його в нашому дослідженні могло б посприяти отриманню більш точних результатів. Перспективною, наприклад, могла б бути інтеграція методу машинного навчання з морфологічним аналізом як основи для класифікації текстових даних.

На наступному етапі роботи було зібрано текстові матеріали з анонімних і журналістських Телеграм-каналів за період чотирьох місяців для проведення порівняльного аналізу. На основі двох генеральних сукупностей публікацій було

укладено дві вибірки, кожна з яких містила по 80 підвбірок по 5000 слів. Аналіз великого об'єму текстових матеріалів дозволив визначити аналізовані показники з високою точністю.

Для обох вибірок було розроблено єдиний алгоритм обробки й обрахунку кількісних і статистичних показників лінгвістичних та екстралінгвістичних ознак. На основі отриманих даних було квантифіковано межі граничних маркерів маніпулятивності, які тепер використовуються для розпізнавання маніпулятивних текстів у Телеграм-боті.

Потенційно проєкт має багато напрямків для подальшого розвитку. Зокрема, до аналізованих у тексті лінгвістичних явищ можна додати різні морфологічні ознаки іменників і дієслів, адже, наприклад, час дієслова може виявитись ефективним маркером для визначення маніпулятивної риторики. Аналогічно може бути розширений спектр екстралінгвістичних критеріїв аналізу: варто здійснити розрахунки для окремих категорій емодзі та чисел, записаних цифрами. Окрім того, у майбутній роботі необхідно знайти спосіб завантажувати публікації з оригінальним форматуванням, що дозволить отримати реальні кількісні показники для покликань, а також – проаналізувати використання різних методів форматування тексту.

Покращення також необхідні в програмному забезпеченні Телеграм-бота. Можливості представленого у цьому дослідженні варіанту є доволі обмеженими. Так, наприклад, Телеграм-бот «Звір – канал» може одночасно перевіряти лише один канал, а обробка матеріалів займає відносно довгий час. Окрім того, для масштабування проєкту важливо перейти з локального на зовнішній хостинг, аби забезпечити постійну доступність бота. Однак, для реалізації цих змін необхідно провести додаткове дослідження, а також інвестувати в проєкт певну суму грошей.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Кулдошина М., Костіков М. Можливості застосування комп'ютерної лінгвістики у військовій справі. *Сучасні методи, інформаційне, програмне та технічне забезпечення систем керування організаційно-технічними та технологічними комплексами* : Міжнар. науково-техн. Internet-конф., м. Київ, 24 листоп. 2023 р. Київ, 2023. С. 101–102. URL: <https://nuft.edu.ua/naukova-diyalnist/naukovi-konferencii> (дата звернення: 09.06.2024).
2. Михайлин І., Основи журналістики : Підручник. 5-те вид. Київ : Центр учб. літ., 2011. 496 с. URL: https://shron1.chtyvo.org.ua/Mykhailyn_Ihor/Osnovy_zhurnlistyky.pdf (дата звернення: 08.06.2024).
3. Опитування USAID-Internews щодо споживання медіа. 2023. 95 с. URL: <https://internews.in.ua/wp-content/uploads/2023/10/Ukrainski-media-stavlennia-ta-dovira-2023r.pdf> (дата звернення: 08.06.2024).
4. Поліщук Ю. Я., Гнатюк С. О., Сейлова Н. А. Мас-медіа як канал маніпулятивного впливу на суспільство. *Ukrainian Scientific Journal of Information Security*. 2015. Т. 21, № 3. URL: <https://doi.org/10.18372/2225-5036.21.9709> (дата звернення: 08.06.2024).
5. Стасула Н. Електронні засоби виявлення вербальної маніпуляції в ЗМІ. *Вісник*. 2011. № 694. С. 90–98. URL: <https://science.lpnu.ua/sites/default/files/journal-paper/2024/feb/33607/vis694komp-nauky-90-98.pdf> (дата звернення: 08.06.2024).
6. Тихомирова Є. Б. Російська інформаційна війна: теоретичні та практично-прикладні аспекти. *Науковий вісник Східноєвропейського національного університету імені Лесі Українки*. 2020. Т. 2, № 406. С. 98–105. URL: <https://evnuir.vnu.edu.ua/handle/123456789/18535> (дата звернення: 08.06.2024).

7. Asch S. E. Effects of group pressure upon the modification and distortion of judgments. *Groups, leadership and men; research in human relations*. 1951. P. 177–190.
8. Asch S. E. Opinions and Social Pressure. *Scientific American*, a division of Nature America, Inc. 1955. P. 31–35.
9. Asch S. E. Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monographs: General and Applied*. 1956. Т. 70, № 9. С. 1–70. URL: <https://doi.org/10.1037/h0093718> (дата звернення: 08.06.2024).
10. Bar-Tal D., Sharvit K., Nets-Zehngut R. Self-Censorship in Contexts of Conflict: Theory and Research. *Springer*, 2018. 301 p.
11. Berger J., Milkman K. L. What Makes Online Content Viral?. *Journal of Marketing Research*. 2012. Т. 49, № 2. С. 192–205. URL: <https://doi.org/10.1509/jmr.10.0353> (дата звернення: 08.06.2024).
12. Brown, D.K., Harlow S., García-Perdomo V., Salaverría R. A new sensation? An international exploration of sensationalism and social media recommendations in online news publications. *Journalism*. 2016. Т. 19, № 11. С. 1497–1516. URL: <https://doi.org/10.1177/1464884916683549> (дата звернення: 08.06.2024).
13. Cavaliere P. From journalistic ethics to fact-checking practices: defining the standards of content governance in the fight against disinformation. *Journal of Media Law*. 2020. Т. 12, № 2. С. 133–165. URL: <https://doi.org/10.1080/17577632.2020.1869486> (дата звернення: 08.06.2024).
14. Deuze M. Journalism and the Web. *Gazette (Leiden, Netherlands)*. 1999. Т. 61, № 5. С. 373–390. URL: <https://doi.org/10.1177/0016549299061005002> (дата звернення: 08.06.2024).
15. Fitzpatrick N. Media Manipulation 2.0: The Impact of Social Media on News, Competition, and Accuracy. *Athens Journal of Mass Media and Communications*.

2018. Т. 4, № 1. С. 45–62. URL: <https://doi.org/10.30958/ajmmc.4.1.3> (дата звернення: 08.06.2024).
16. Gabielkov M., Ramachandran A., Chaintreau A., Legou A. Social Clicks: What and Who Gets Read on Twitter? *ACM SIGMETRICS Performance Evaluation Review*. 2016. Т. 44, № 1. С. 179–192. URL: <https://doi.org/10.1145/2964791.2901462> (дата звернення: 08.06.2024).
17. Kaplan R. L. Yellow journalism. *The international encyclopedia of communication*. 2008. P. 5369–5370.
18. Kapusta J., Drlik M., Munk M. Using of n-grams from morphological tags for fake news classification. *PeerJ Computer Science*. 2021. Vol. 7. P. e624. URL: <https://doi.org/10.7717/peerj-cs.624> (дата звернення: 08.06.2024).
19. Kapusta, J., Hájek, P., Munk, M., & Benko, L. Comparison of fake and real news based on morphological analysis. *Procedia Computer Science*. 2020. Vol. 171. P. 2285–2293. URL: <https://doi.org/10.1016/j.procs.2020.04.247> (дата звернення: 08.06.2024).
20. Keeble R. Ethics for journalists. 2-ге вид. *Taylor & Francis Group*, 2009. 326 с. URL: https://library.uniteddiversity.coop/Media_and_Free_Culture/Ethics_for_journalists-Richard_Keeble.pdf (дата звернення: 08.06.2024).
21. Konstankevych, I., Kostusiak, N., Shulska, N., Stanislav, O., Yelova, T., Kauza, I. Media Manipulation as a Tool of Information Warfare: Typology Signs, Language Markers, Fact Checking Methods. *Ad Alta*. 2022. Vol. 2, no. 12. P. 224–230. URL: <https://evnuir.vnu.edu.ua/handle/123456789/22040> (дата звернення: 08.06.2024).
22. Mahyoob M., Algaraady J., Alrahaili M. Linguistic-Based Detection of Fake News in Social Media. *International Journal of English Linguistics*. 2020. Vol. 11, no. 1. P. 99. URL: <https://doi.org/10.5539/ijel.v11n1p99> (дата звернення: 08.06.2024).

23. Marwick A. E., Lewis R. Media manipulation and disinformation online. *Data & Society Research Institute*, 2017. 104 с. URL: <https://apo.org.au/sites/default/files/resource-files/2017-05/apo-nid135936.pdf> (дата звернення: 08.06.2024).
24. Semenovych I. M., Alieinyk V. H. Russian aggression against Ukraine in 2022–2023: specifics of mass communication. *Актуальні проблеми філософії та соціології*. 2022. № 39. С. 136–141. URL: <https://doi.org/10.32782/apfs.v039.2022.23> (дата звернення: 08.06.2024).
25. Stoick B., Snell N., Straub J. Fake news identification: a comparison of parts-of-speech and N-grams with neural networks. *Big Data: Learning, Analytics, and Applications*. 2019.
26. Vosoughi S., Roy D., Aral S. The spread of true and false news online. *Science*. 2018. Т. 359, № 6380. С. 1146–1151. URL: <https://doi.org/10.1126/science.aap9559> (дата звернення: 08.06.2024).

ЕЛЕКТРОННІ ДЖЕРЕЛА

27. Анонімність та російські гроші. Хто фінансує українські телеграм-канали? (оновлено). *Інститут масової інформації*. URL: <https://imi.org.ua/monitorings/anonymist-ta-rosijski-groschi-hto-finansuye-ukrayinski-telegram-kanaly-i51944> (дата звернення: 08.06.2024).
28. Білий список: 11 медіа, що стали найякіснішими. *Інститут масової інформації*. URL: <https://imi.org.ua/news/bilyj-spysok-11-media-shho-staly-najyakisnishymy-i60964> (дата звернення: 08.06.2024).
29. Головна. *StopFake*. URL: <https://www.stopfake.org/> (дата звернення: 08.06.2024).
30. Документи. *НСЖУ*. URL: <https://nsju.org/spilka/dokumenty/#1642060895505-0c2d8311-e930> (дата звернення: 08.06.2024).
31. Дріт. *Telegram*. URL: <https://t.me/hrinovyny> (дата звернення: 08.06.2024).
32. Інформаційна гігієна для людей та компаній | Як не стати овочем. *oksanamoroz*. URL: <https://www.oksanamoroz.com/> (дата звернення: 08.06.2024).
33. Майже ЗМІ. Як Telegram маніпулює аудиторією. *Інститут масової інформації*. URL: <https://imi.org.ua/monitorings/majzhe-zmi-yak-telegram-manipulyuye-audytoriye-yu-i49222> (дата звернення: 08.06.2024).
34. Радіо Свобода. *Telegram*. URL: https://t.me/svoboda_radio (дата звернення: 08.06.2024).
35. Радник. *Telegram*. URL: <https://t.me/+PsA00CQQwHthODQ6> (дата звернення: 08.06.2024).
36. Труханули телеграм | Розслідування. *NGL.media*. URL: <https://ngl.media/2023/09/05/truhanuli-telegram/> (дата звернення: 08.06.2024).

37. Український тиждень. *Telegram*. URL: <https://t.me/tyzhdenUA> (дата звернення: 08.06.2024).
38. 100 російських ТГ-каналів, які мімікують під українські. *SPRAVDI*. URL: <https://spravdi.gov.ua/spilna-zayava-z-pryvodu-zahystu-informacijnogo-prostoru-ukrayiny-vid-rosijskyh-vorozhyh-telegram-kanaliv/> (дата звернення: 08.06.2024).
39. 3.12.4 documentation. *Python*. URL: <https://docs.python.org/3/> (дата звернення: 08.06.2024).
40. Asyncio – asynchronous I/O. *Python documentation*. URL: <https://docs.python.org/3/library/asyncio.html> (дата звернення: 08.06.2024).
41. BotFather. *Telegram*. URL: <https://telegram.me/BotFather> (дата звернення: 08.06.2024).
42. Collections – Container datatypes. *Python documentation*. URL: <https://docs.python.org/3/library/collections.html> (дата звернення: 08.06.2024).
43. Csv – CSV file reading and writing. *Python documentation*. URL: <https://docs.python.org/3/library/csv.html> (дата звернення: 08.06.2024).
44. Datetime – Basic date and time types. *Python documentation*. URL: <https://docs.python.org/3/library/datetime.html> (дата звернення: 08.06.2024).
45. DeepStateUA. *Telegram*. URL: <https://t.me/DeepStateUA> (дата звернення: 08.06.2024).
46. detector.media. Детектор медіа. *detector.media*. URL: <https://detector.media/> (дата звернення: 08.06.2024).
47. Емоїї. *PyPI*. URL: <https://pypi.org/project/emoji/> (дата звернення: 08.06.2024).
48. FactCheck.org. FactCheck.org. URL: <https://www.factcheck.org/> (дата звернення: 08.06.2024).
49. Google Colab. *Google Colab*. URL: <https://colab.research.google.com/> (дата звернення: 08.06.2024).

50. Io – Core tools for working with streams. *Python documentation*. URL: <https://docs.python.org/3/library/io.html> (дата звернення: 08.06.2024).
51. Math – Mathematical functions. *Python documentation*. URL: <https://docs.python.org/3/library/math.html> (дата звернення: 08.06.2024).
52. Matplotlib – Visualization with Python. *Matplotlib*. URL: <https://matplotlib.org/> (дата звернення: 08.06.2024).
53. Microsoft. Visual Studio Code - Code Editing. Redefined. *Visual Studio Code*. URL: <https://code.visualstudio.com/> (дата звернення: 08.06.2024).
54. NLTK :: Natural Language Toolkit. *NLTK*. URL: <https://www.nltk.org/> (дата звернення: 08.06.2024).
55. Pandas.DataFrame – pandas 2.2.2 documentation. *pandas - Python Data Analysis Library*. URL: <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.html> (дата звернення: 08.06.2024).
56. Pandas. *PyPI*. URL: <https://pypi.org/project/pandas/> (дата звернення: 08.06.2024).
57. Pytz. *PyPI*. URL: <https://pypi.org/project/pytz/> (дата звернення: 08.06.2024).
58. Re – Regular expression operations. *Python documentation*. URL: <https://docs.python.org/3/library/re.html> (дата звернення: 08.06.2024).
59. SpaCy · industrial-strength natural language processing in python. *spaCy*. URL: <https://spacy.io/> (дата звернення: 08.06.2024).
60. Telebot. *PyPI*. URL: <https://pypi.org/project/telebot/> (дата звернення: 08.06.2024).
61. Telethon. *PyPI*. URL: <https://pypi.org/project/Telethon/> (дата звернення: 08.06.2024).
62. The definitive fact-checking site and reference source for urban legends, folklore, myths, rumors, and misinformation. *Snopes.com*. URL: <https://www.snopes.com/> (дата звернення: 08.06.2024).

63. Transparent reliability ratings for news and information sources. *NewsGuard*.
URL: <https://www.newsguardtech.com/> (дата звернення: 08.06.2024).
64. TreeTagger. Centrum für Informations- und Sprachverarbeitung - LMU Munich.
URL: <https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/> (дата звернення: 08.06.2024).
65. W3Schools.com. *W3Schools Online Web Tutorials*. URL:
https://www.w3schools.com/python/ref_string_isupper.asp (дата звернення:
08.06.2024).
66. ZN.UA. *Telegram*. URL: https://t.me/znuu_live (дата звернення: 08.06.2024).

СПИСОК ДЖЕРЕЛ ТЕКСТОВОГО МАТЕРІАЛУ

- Адвокат Мілер. *Telegram*. URL: <https://t.me/+YTGTWT9qNlXhNjMy> (дата звернення: 08.06.2024).
- Батальйон Монако. *Telegram*. URL: <https://t.me/+pLuXtjjNmz83ODdi> (дата звернення: 08.06.2024).
- Бабель. *Telegram*. URL: https://t.me/hromadske_ua (дата звернення: 08.06.2024).
- Доброго вечора, ми з України. *Telegram*. URL: https://t.me/from_ukraine_we (дата звернення: 08.06.2024).
- Еспресо. *Telegram*. URL: <https://t.me/espresotb> (дата звернення: 08.06.2024).
- Громадське. *Telegram*. URL: https://t.me/hromadske_ua (дата звернення: 08.06.2024).
- Суспільне. *Telegram*. URL: <https://t.me/suspilnews> (дата звернення: 08.06.2024).
- Труха Україна. *Telegram*. URL: <https://t.me/truexanewsua> (дата звернення: 08.06.2024).
- Українська правда. *Telegram*. URL: https://t.me/ukrpravda_news (дата звернення: 08.06.2024).
- Харків 1654. *Telegram*. URL: https://t.me/kharkiv_1654 (дата звернення: 08.06.2024).

ДОДАТКИ

Додаток 1. Програмний код

1. **Посилання на повний код алгоритму обчислення кількісних і статистичних показників для маркерів маніпуляції**

Усі модулі за посиланням:

https://colab.research.google.com/drive/1S8i98__zxwHe2VNlzGS0jbzzWZYLYf_bu?usp=sharing

2. **Посилання на повний код Телеграм-бота**

Усі файли за посиланням:

https://github.com/msklyar13/zvirchannel_bot

Додаток 2. Текстові матеріали

Усі файли з використаними текстовими матеріалами наявні за посиланням:

https://drive.google.com/drive/folders/11zaAHi1GlqmJ3n565NP_LmkaAk-6wXXt?usp=sharing

Додаток 3. Результати статистичних обрахунків показників лінгвістичних і екстралінгвістичних ознак

1. Таблиці з середніми частотами та довірчими інтервалами для проаналізованих лінгвістичних і екстралінгвістичних ознак:

○ покликання

	Середня частота	Довірчий інтервал σ	Довірчий інтервал 2σ	Довірчий інтервал 3σ
Внутрішні				
Маніпулятивні джерела	5.81	-0.87 - 12.49	-7.55 - 19.17	-14.23 - 25.85
Надійні джерела	0.01	-0.1 - 0.12	-0.21 - 0.23	-0.32 - 0.34
Зовнішні				
Маніпулятивні джерела	2.76	-1.38 - 6.9	-5.52 - 11.04	-9.66 - 15.18
Надійні джерела	1	-0.6 - 2.6	-2.2 - 4.2	-3.8 - 5.8

○ емодзі

	Середня частота	Довірчий інтервал σ	Довірчий інтервал 2σ	Довірчий інтервал 3σ
Маніпулятивні джерела	303.51	260.2 - 346.82	216.89 - 390.13	173.58 - 433.44
Надійні джерела	123.47	47.65 - 199.29	-28.17 - 275.11	-103.99 - 350.93

○ слова у верхньому регістрі

	Середня частота	Довірчий інтервал σ	Довірчий інтервал 2σ	Довірчий інтервал 3σ
Маніпулятивні джерела	175.21	121.65 - 228.77	68.09 - 282.33	14.53 - 335.89
Надійні джерела	101.51	76.38 - 126.64	51.25 - 151.77	26.12 - 176.9

○ власні назви

	Середня частота	Довірчий інтервал σ	Довірчий інтервал 2σ	Довірчий інтервал 3σ
Маніпулятивні джерела	359.24	282.69 - 435.79	206.14 - 512.34	129.59 - 588.89
Надійні джерела	421.32	380.95 - 461.69	340.58 - 502.06	300.21 - 542.43

○ частини мови

	Середня частота	Довірчий інтервал σ	Довірчий інтервал 2σ	Довірчий інтервал 3σ
Іменники				
Маніпулятивні джерела	1688.15	1622.04 - 1754.26	1555.93 - 1820.37	1489.82 - 1886.48
Надійні джерела	1711	1653.86 - 1768.14	1596.72 - 1825.28	1539.58 - 1882.42
Дієслова				
Маніпулятивні джерела	629.99	581.65 - 678.33	533.31 - 726.67	484.97 - 775.01
Надійні джерела	635.04	604.6 - 665.48	574.16 - 695.92	543.72 - 726.36
Прикметники				
Маніпулятивні джерела	500.01	453.57 - 546.45	407.13 - 592.89	360.69 - 639.33
Надійні джерела	591.05	545.2 - 636.9	499.35 - 682.75	453.5 - 728.6
Прислівники				
Маніпулятивні джерела	279.39	241.8 - 316.98	204.21 - 354.57	166.62 - 392.16
Надійні джерела	213.94	190.77 - 237.11	167.6 - 260.28	144.43 - 283.45
Числівники				
Маніпулятивні джерела	19.48	12.72 - 26.24	5.96 - 33.0	-0.8 - 39.76

Надійні джерела	41.7	22.81 - 60.59	3.92 - 79.48	-14.97 - 98.37
Займенники				
Маніпулятивні джерела	164.66	134.93 - 194.39	105.2 - 224.12	75.47 - 253.85
Надійні джерела	121.41	98.85 - 143.97	76.29 - 166.53	53.73 - 189.09
Сполучники сурядності				
Маніпулятивні джерела	169.47	148.8 - 190.14	128.13 - 210.81	107.46 - 231.48
Надійні джерела	166.69	145.3 - 188.08	123.91 - 209.47	102.52 - 230.86
Сполучники підрядності				
Маніпулятивні джерела	78.83	56.26 - 101.4	33.69 - 123.97	11.12 - 146.54
Надійні джерела	84.51	64.91 - 104.11	45.31 - 123.71	25.71 - 143.31
Частки				
Маніпулятивні джерела	118.36	91.96 - 144.76	65.56 - 171.16	39.16 - 197.56
Надійні джерела	63.92	48.47 - 79.37	33.02 - 94.82	17.57 - 110.27

○ особові займенники

	Середня частота	Довірчий інтервал σ	Довірчий інтервал 2σ	Довірчий інтервал 3σ
Перша особа				
Маніпулятивні джерела	27.84	18.7 - 36.98	9.56 - 46.12	0.42 - 55.26
Надійні джерела	15.28	6.32 - 24.24	-2.64 - 33.2	-11.6 - 42.16
Друга особа				
Маніпулятивні джерела	18.95	13.36 - 24.54	7.77 - 30.13	2.18 - 35.72
Надійні джерела	2.46	-0.48 - 5.4	-3.42 - 8.34	-6.36 - 11.28
Третя особа				

Маніпулятивні джерела	50.45	36.18 - 64.72	21.91 - 78.99	7.64 - 93.26
Надійні джерела	61.2	49.32 - 73.08	37.44 - 84.96	25.56 - 96.84

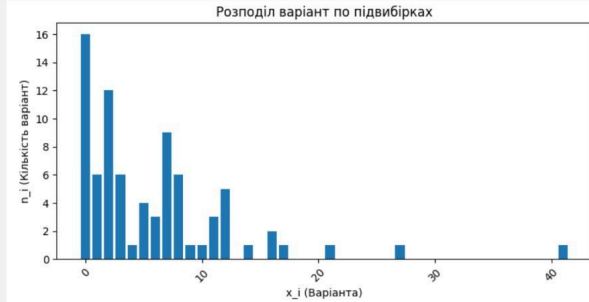
○ порівняльні ступені прикметників

	Середня частота	Довірчий інтервал σ	Довірчий інтервал 2σ	Довірчий інтервал 3σ
Вищий ступінь порівняння				
Маніпулятивні джерела	3.96	1.62 - 6.3	-0.72 - 8.64	-3.06 - 10.98
Надійні джерела	3.85	1.28 - 6.42	-1.29 - 8.99	-3.86 - 11.56
Найвищий ступінь порівняння				
Маніпулятивні джерела	9.66	4.8 - 14.52	-0.06 - 19.38	-4.92 - 24.24
Надійні джерела	6.41	3.64 - 9.18	0.87 - 11.95	-1.9 - 14.72

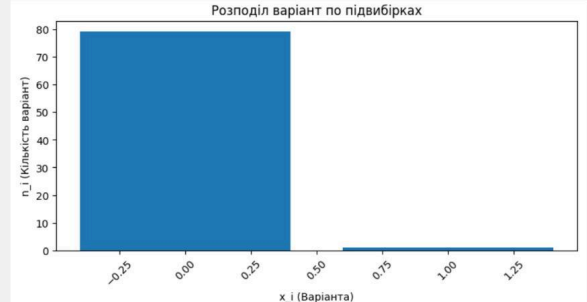
2. Стовпчикові діаграми варіативних рядів проаналізованих лінгвістичних і екстралінгвістичних ознак

Внутрішні покликання

Анонімні джерела

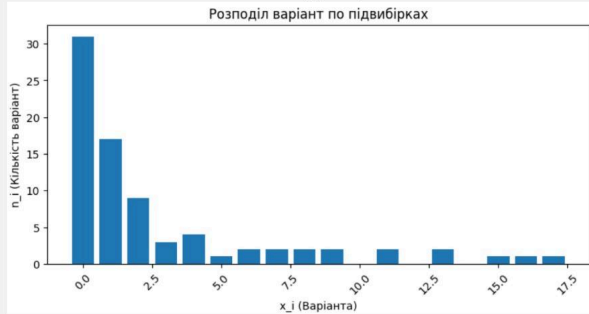


Журналістські джерела

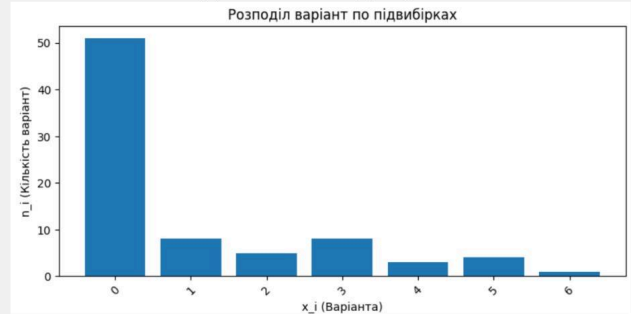


Зовнішні покликання

Анонімні джерела

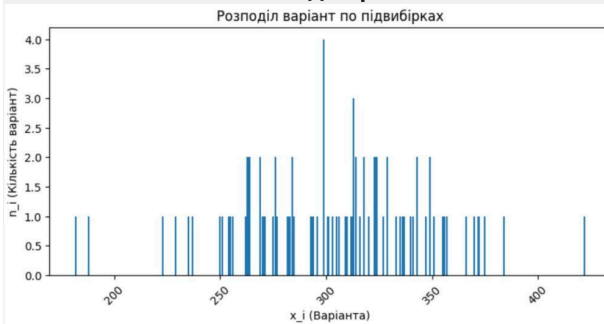


Журналістські джерела

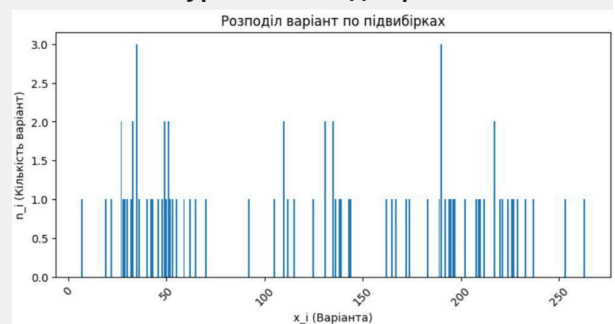


Емодзі

Анонімні джерела

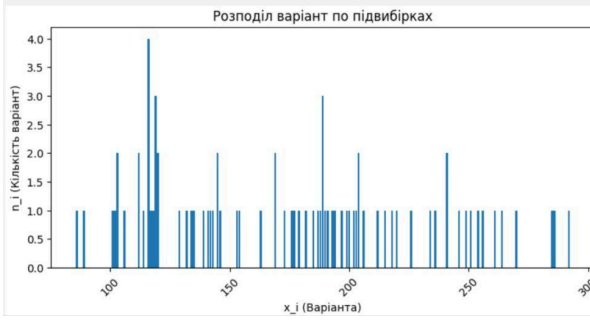


Журналістські джерела

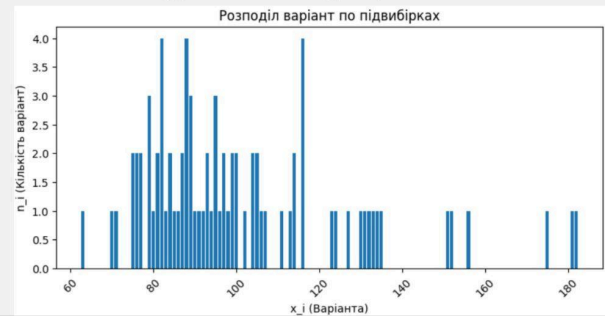


Слова у верхньому регістрі

Анонімні джерела

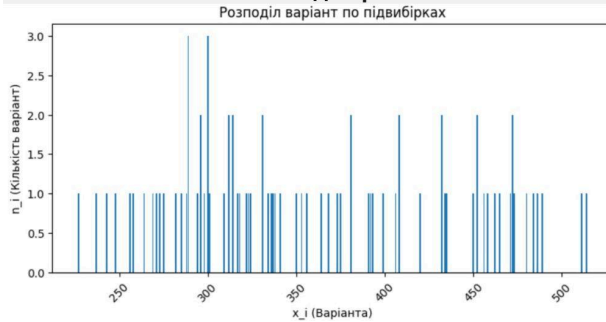


Журналістські джерела

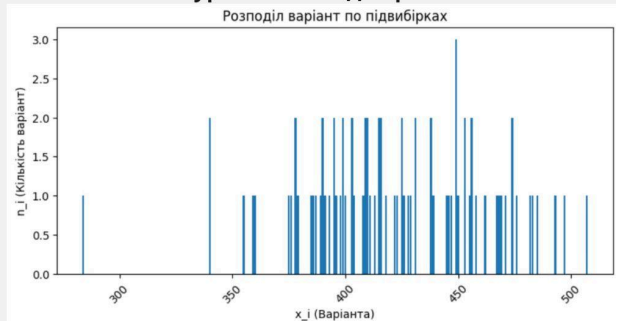


Власні назви

Анонімні джерела

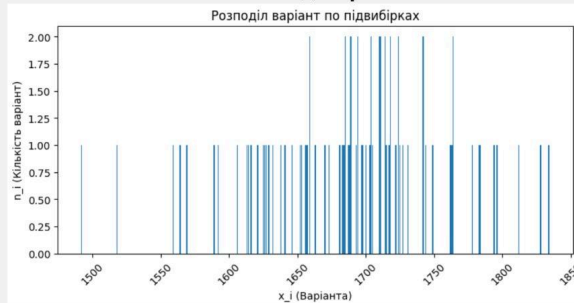


Журналістські джерела

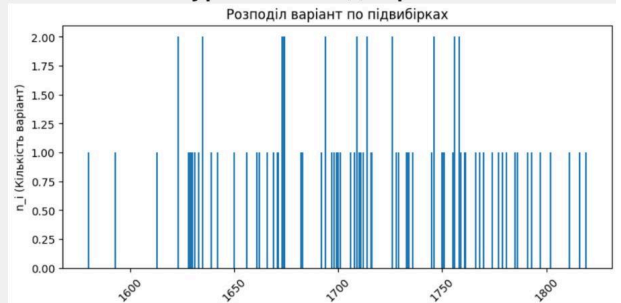


Іменники

Анонімні джерела

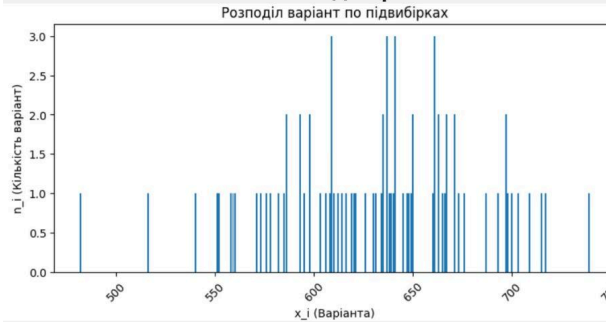


Журналістські джерела

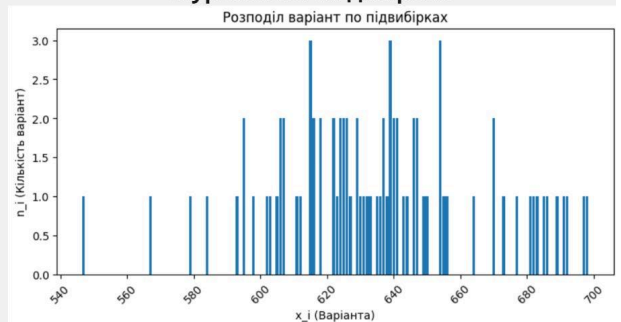


Дієслова

Анонімні джерела

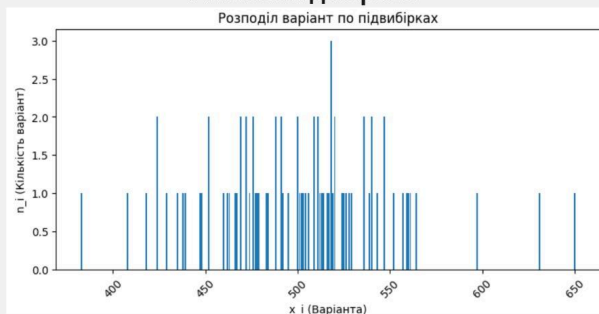


Журналістські джерела

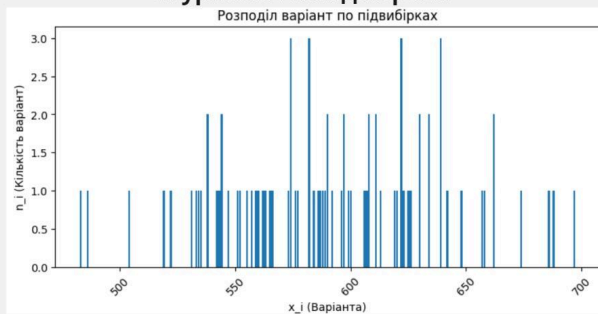


Прикметники

Анонімні джерела

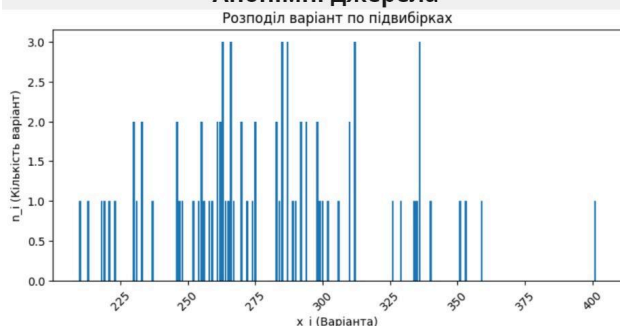


Журналістські джерела

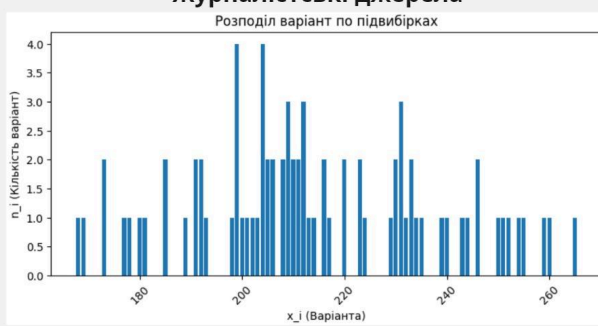


Прислівники

Анонімні джерела

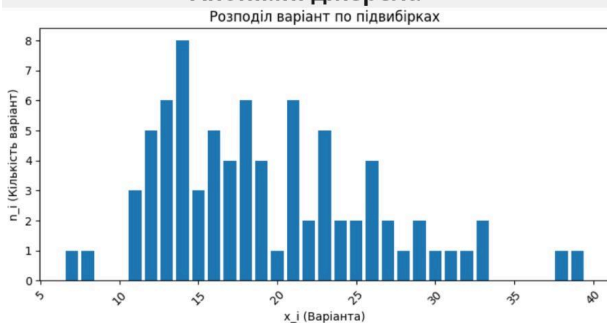


Журналістські джерела

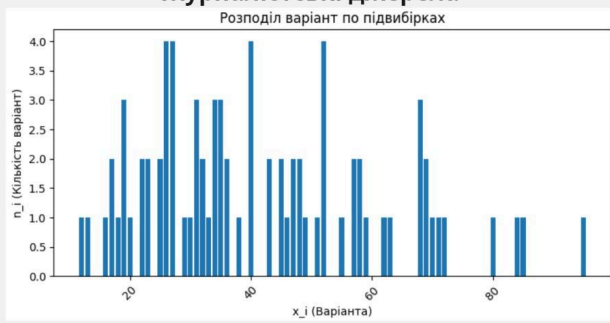


Числівники

Анонімні джерела

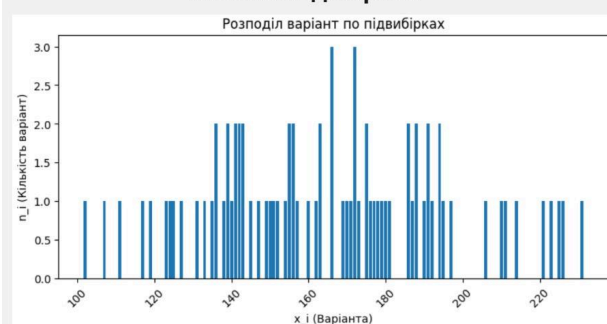


Журналістські джерела

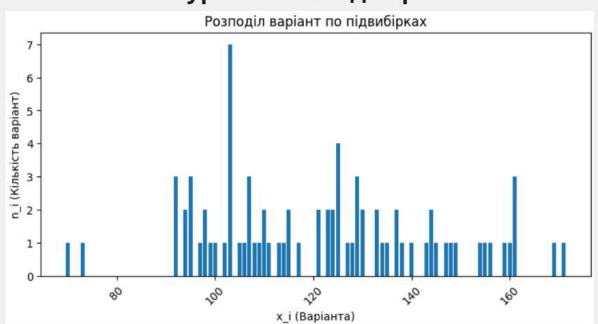


Займенники

Анонімні джерела

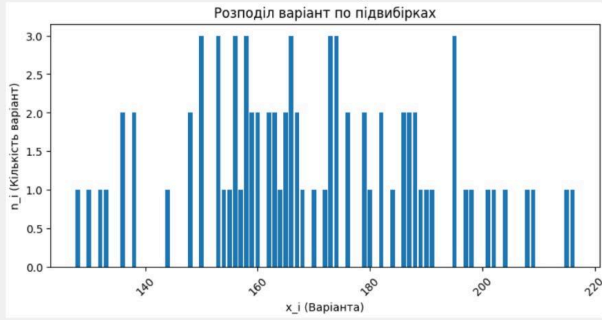


Журналістські джерела

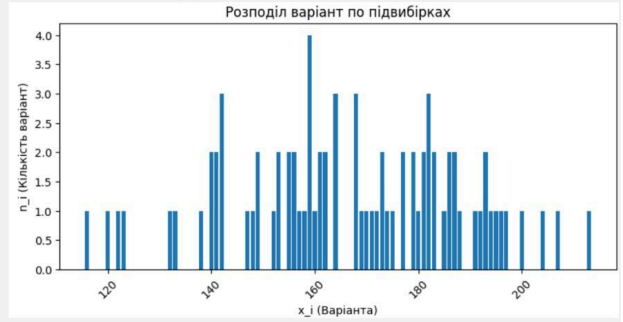


Сполучники сурядності

Анонімні джерела

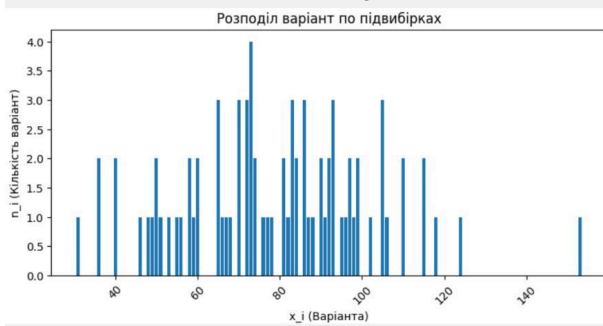


Журналістські джерела

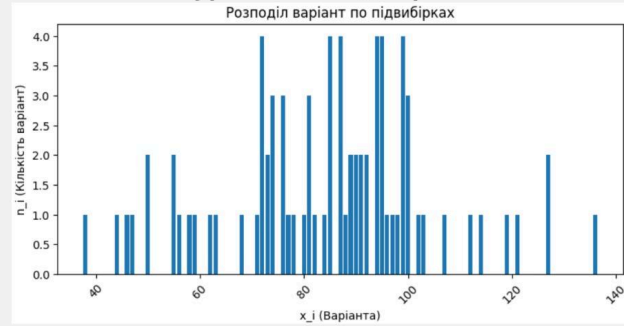


Сполучники підрядності

Анонімні джерела

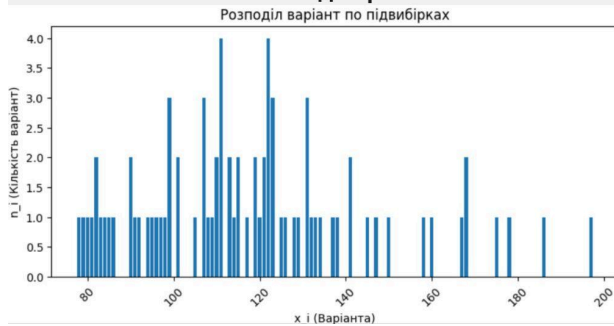


Журналістські джерела

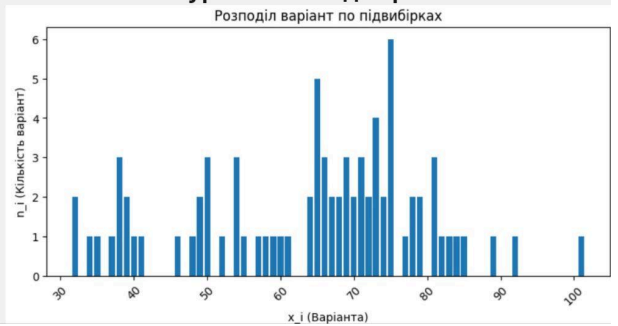


Частки

Анонімні джерела



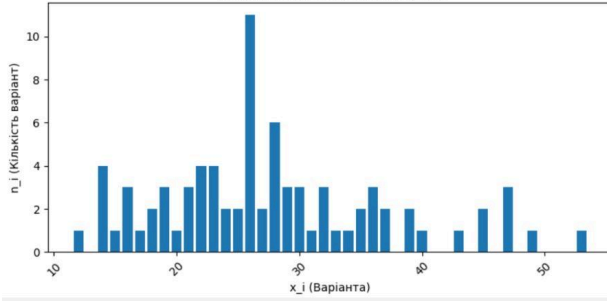
Журналістські джерела



Займенники першої особи

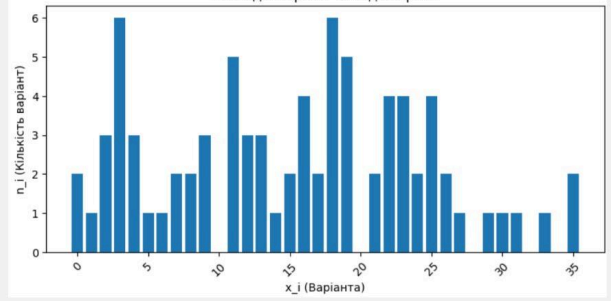
Анонімні джерела

Розподіл варіант по підвбірках



Журналістські джерела

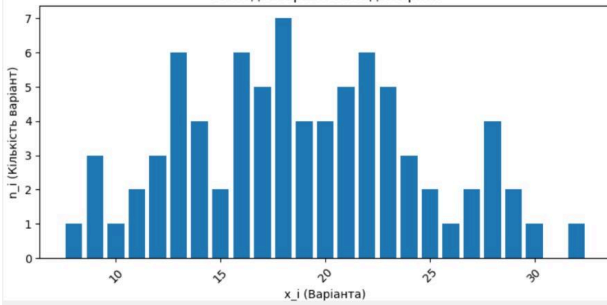
Розподіл варіант по підвбірках



Займенники другої особи

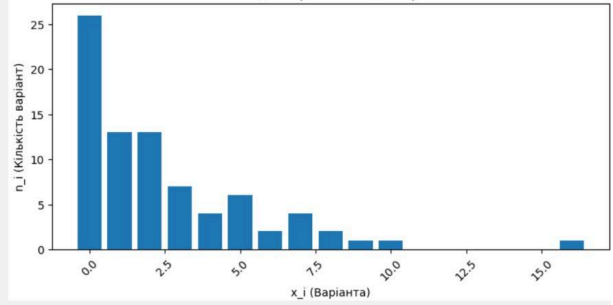
Анонімні джерела

Розподіл варіант по підвбірках



Журналістські джерела

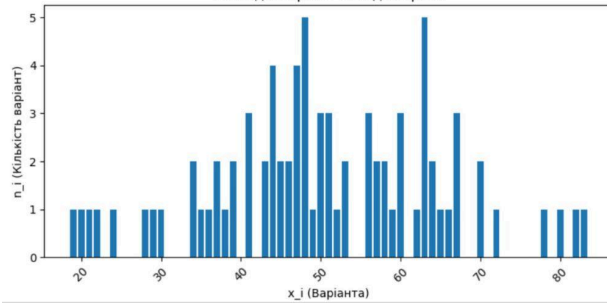
Розподіл варіант по підвбірках



Займенники третьої особи

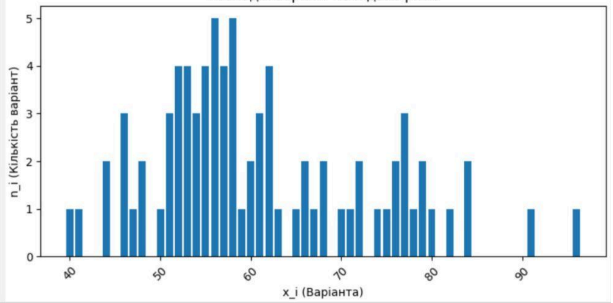
Анонімні джерела

Розподіл варіант по підвбірках



Журналістські джерела

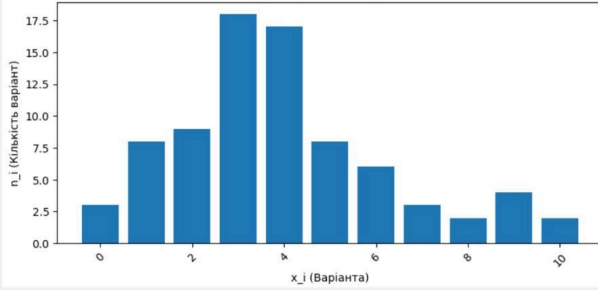
Розподіл варіант по підвбірках



Вищий ступінь порівняння прикметників

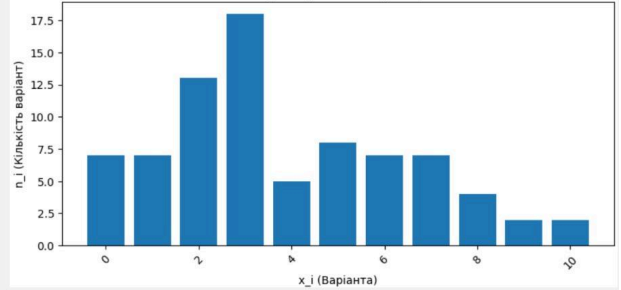
Анонімні джерела

Розподіл варіант по підвбірках



Журналістські джерела

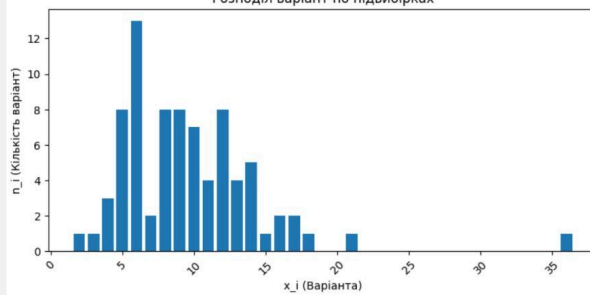
Розподіл варіант по підвбірках



Найвищий ступінь порівняння прикметників

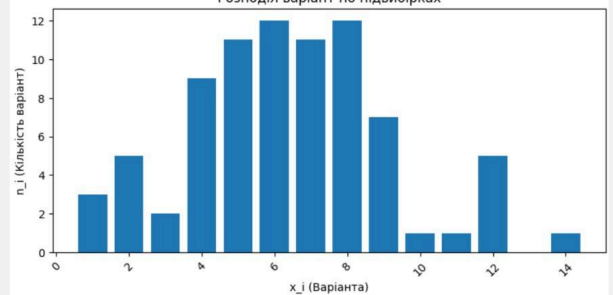
Анонімні джерела

Розподіл варіант по підвбірках



Журналістські джерела

Розподіл варіант по підвбірках



Додаток 4. Вміст файлу «markers.csv»

parameter	grey_min	grey_max
int_links	-0.21	0.23
ext_links	-2.2	4.2
emoji		275.11
capitalization		151.77
noun	1596.72	
propn	340.58	502.06
verb	574.16	695.92
adj	499.35	
adv		260.28
num		
pron		166.53
cconj		209.47
sconj	45.31	123.71
part		94.82
pers1		33.2
pers2		8.34
pers3	37.44	
comp		
sup	0.87	11.95