

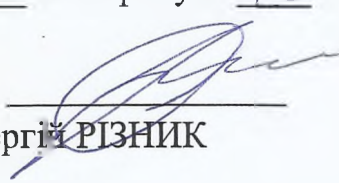
Міністерство освіти і науки України
Київський національний університет імені Тараса Шевченка
Навчально-науковий інститут філології
Кафедра української мови та прикладної лінгвістики

**АВТОМАТИЗОВАНИЙ АНАЛІЗ ВЕРБАЛЬНИХ СТРАТЕГІЙ
ВПЛИВУ В УКРАЇНСЬКОМУ НОВИННОМУ СЕРЕДОВИЩІ**

Кваліфікаційна робота
на здобуття ОС «бакалавр»
студентки IV курсу
галузі знань 03 «Гуманітарні науки»,
спеціальність 035 «Філологія»
спеціалізації 035.10 «Прикладна
лінгвістика», ОПП «Прикладна
(комп'ютерна) лінгвістика
та англійська мова»
Анастасії Павлівни МАНЗІЙ

Науковий керівник:
к.т.н., доц. **Микола КОСТИКОВ**

«Допущено до захисту»
Протокол засідання кафедри
української мови та прикладної лінгвістики
від « 5 » 06 2026 року № 16

завідувач кафедри 
к.філол.н., доц. **Сергій РІЗНИК**

КИЇВ - 2026

АНОТАЦІЯ

Манзій А. П. Автоматизований аналіз вербальних стратегій впливу в українському новинному середовищі. Кваліфікаційна робота бакалавра. Київський національний університет імені Тараса Шевченка, Київ, 2026.

Дослідження присвячене вивченню вербальних стратегій впливу в україномовному новинному Telegram-дискурсі воєнного часу та розробці системи їх автоматизованої класифікації засобами машинного навчання. Актуальність роботи зумовлена домінуючою роллю Telegram як новинної платформи в умовах повномасштабного вторгнення та стрімким зростанням обсягів контенту, що унеможлиблює його ручний аналіз. Об'єктом дослідження є новинна комунікація в українському сегменті Telegram, предметом – вербальні стратегії впливу та їх автоматизована класифікація.

Робота складається з трьох розділів. У теоретичному систематизовано підходи до класифікації вербальних стратегій та охарактеризовано Telegram як середовище їх реалізації. У двох практичних описано збір корпусу з 9939 повідомлень, ручну розмітку датасету та побудову класифікатора. Методологічну основу становлять контент-аналіз, дискурс-аналіз та методи машинного навчання. Класифікатор побудовано на основі fine-tuning моделі bert-base-multilingual-cased із багатомітковою класифікацією та зваженою функцією втрат; досягнуто F1-score 0.75. Порівняльний аналіз встановив кореляцію між типом каналу та його стратегічним профілем. Результати можуть бути застосовані для моніторингу медіапростору та підвищення медіаграмотності.

Ключові слова: вербальні стратегії впливу, медіадискурс, Telegram, машинне навчання, BERT, класифікація тексту, новинна комунікація, воєнний час.

ABSTRACT

Manzii A. P. Automated Analysis of Verbal Influence Strategies in the Ukrainian News Environment. Bachelor's qualification thesis. Taras Shevchenko National University of Kyiv, Kyiv, 2026.

The study focuses on verbal influence strategies in Ukrainian-language wartime Telegram news discourse and the development of an automated classification system using machine learning. The relevance of the research is determined by Telegram's dominant role as a news platform during the full-scale invasion and the rapid growth of content volumes that makes manual analysis impractical. The object of the study is news communication in the Ukrainian segment of Telegram; the subject is verbal influence strategies and their automated classification.

The thesis consists of three chapters. The theoretical chapter systematizes approaches to classifying verbal strategies and characterizes Telegram as an environment for their realization. The two practical chapters describe the collection of a corpus of 9939 messages, manual annotation of the training dataset, and construction of the classifier. The methodological framework combines content analysis, discourse analysis, and machine learning methods. The classifier was built by fine-tuning the bert-base-multilingual-cased model with multi-label classification and a weighted loss function, achieving an F1-score of 0.75. Comparative analysis established a correlation between channel type and its strategic profile. The results can be applied to media monitoring and improving media literacy.

Keywords: verbal influence strategies, media discourse, Telegram, machine learning, BERT, text classification, news communication, wartime.

ЗМІСТ

ВСТУП.....	7
РОЗДІЛ 1. ТЕОРЕТИЧНІ ЗАСАДИ ДОСЛІДЖЕННЯ МОВНИХ СТРАТЕГІЙ У МЕДІА.....	10
1.1. Поняття мовних стратегій: класифікація, функції та підходи до визначення.....	10
1.2. Особливості цифрового медіапростору та Telegram як платформа реалізації стратегій впливу.....	13
1.3. Маніпуляція в медіа: техніки, цілі та адаптація до аудиторії.....	17
Висновки до першого розділу.....	21
РОЗДІЛ 2. ФОРМУВАННЯ КОРПУСУ ДАНИХ ТА МЕТОДОЛОГІЯ ДОСЛІДЖЕННЯ.....	22
2.1. Методи дослідження та огляд матеріалів для аналізу.....	22
2.2. Програмна реалізація системи автоматизованого збору та попередньої обробки даних.....	25
Висновки до другого розділу.....	34
РОЗДІЛ 3. ЕКСПЕРТНА РОЗМІТКА ТА АВТОМАТИЗОВАНА КЛАСИФІКАЦІЯ МЕТОДАМИ МАШИННОГО НАВЧАННЯ.....	36
3.1. Принципи лінгвістичного анотування та формування тренувального датасету.....	36
3.2. Архітектура, налаштування та навчання моделі класифікації вербальних стратегій.....	42
3.3. Автоматизована розмітка корпусу.....	49
3.4. Оцінка ефективності автоматизованого аналізу та інтерпретація результатів.....	53
Висновки до третього розділу.....	57
ВИСНОВКИ.....	59
СПИСОК ПОСИЛАНЬ.....	61
ДОДАТКИ.....	66

ВСТУП

Сучасне інформаційне суспільство характеризується стрімким зростанням обсягів комунікації, де ключову роль на українському медіаринку відіграє платформа Telegram. В умовах повномасштабного вторгнення та інформаційного протистояння телеграм-канали стали основним джерелом оперативної інформації для мільйонів українців, що значно підвищує їхній вплив на формування громадської думки, емоційний стан та соціальні орієнтири аудиторії. Зазначені обставини зумовлюють особливу **актуальність** вивчення механізмів мовного впливу в цьому середовищі – як для наукового осмислення медіакомунікацій воєнного часу, так і для практичних потреб інформаційної безпеки та медіаграмотності.

Одним із ключових інструментів такого впливу є вербальні стратегії – сукупність мовних засобів і прийомів, які дозволяють авторам медіаконтенту досягати певних комунікативних цілей. Їх використання набуває особливої ваги в умовах війни, політичної нестабільності та інформаційного протистояння. Водночас стрімке зростання обсягів новинного контенту робить ручний аналіз таких стратегій практично неможливим, що зумовлює необхідність автоматизованих підходів до їх виявлення та класифікації.

Мета бакалаврської роботи – розробити систему автоматизованого аналізу вербальних стратегій впливу в українському новинному Telegram-середовищі та дослідити закономірності їх використання на матеріалі реальних даних.

Основні завдання дослідження:

- з'ясувати сутність вербальних стратегій впливу та схарактеризувати їх основні типи в медіадискурсі;
- розглянути особливості реалізації вербальних стратегій у Telegram-каналах новинного характеру;
- сформувати корпус повідомлень з обраних каналів та здійснити часткову ручну розмітку вербальних стратегій;

- побудувати модель машинного навчання для автоматичної класифікації вербальних стратегій;
- проаналізувати результати автоматизованої розмітки та виявити закономірності у використанні стратегій впливу в різних типах каналів.

Об'єкт дослідження – новинна комунікація в українському сегменті Telegram.

Предмет дослідження – вербальні стратегії впливу в україномовних новинних повідомленнях та їх автоматизована класифікація засобами машинного навчання.

Матеріалом дослідження стали 9952 повідомлень з п'яти україномовних Telegram-каналів різних типів (Труха ⚡ Україна, Реальна Війна | Україна | Новини, Телеграмна служба новин - Україна, Лачен пише, Zelenskiy / Official), зібраних за період (грудень 2023 – січень 2024). Канали відібрані за типологічним принципом і охоплюють анонімні агрегатори, авторський канал та офіційний державний канал, що дозволяє порівняти вербальні стратегії різних суб'єктів медіапростору.

Методи дослідження охоплюють:

- аналітичний метод – для теоретичного осмислення поняття вербальної стратегії та медіаконтексту;
- контент-аналіз – для системного виявлення вербальних засобів у новинних текстах;
- дискурс-аналіз – для вивчення контексту і прихованого впливу повідомлень на сприйняття аудиторією;
- методи машинного навчання – для автоматизованої класифікації вербальних стратегій на основі частково розміченого корпусу даних.

Новизна дослідження полягає у поєднанні лінгвістичного аналізу вербальних стратегій впливу з методами машинного навчання на матеріалі україномовного Telegram-середовища воєнного часу. На відміну від попередніх досліджень, що спираються виключно на ручний аналіз обмеженої вибірки,

запропонований підхід дозволяє масштабувати аналіз на більший корпус даних і виявляти системні закономірності у використанні стратегій.

Практичне значення роботи полягає в можливості використання розробленої системи для моніторингу медіапростору, виявлення маніпулятивних стратегій у новинному контенті та підвищення медіаграмотності аудиторії. Отримані результати можуть бути застосовані дослідниками медіакомунікацій, журналістами-аналітиками, а також у сфері інформаційної безпеки.

Структура роботи: бакалаврська робота складається зі вступу, трьох основних розділів (теоретичного та двох практичних), висновків, списку джерел і додатків, що включають базу даних повідомлень із результатами автоматизованої розмітки.

РОЗДІЛ 1. ТЕОРЕТИЧНІ ЗАСАДИ ДОСЛІДЖЕННЯ МОВНИХ СТРАТЕГІЙ У МЕДІА

У сучасному інформаційному суспільстві засоби масової інформації впливають на життя кожної людини, великою мірою визначають характер поведінки та життєдіяльності людей (Костирко, Корольова, Жигалкіна, 2019, с.3). Вони відіграють ключову роль у формуванні громадської думки та впливі на поведінку аудиторії. Це досягається завдяки використанню різноманітних мовних стратегій, які забезпечують ефективну комунікацію та маніпуляцію інформацією. У цьому розділі розглянуто поняття мовних стратегій, їх класифікацію та функції, які вони виконують у медіа. Окрему увагу приділено особливостям цифрового інформаційного простору, зокрема в соціальних мережах, де такі стратегії найбільше прослідковуються.

1.1. Поняття мовних стратегій: класифікація, функції та підходи до визначення

У теорії мовної комунікації стратегію мовленнєвого спілкування описують як свідомий та цілеспрямований процес вибору найбільш оптимальних мовних засобів та прийомів для досягнення певної мети спілкування, від широких організаційних та ідеологічних цілей і до більш практичних, пов'язаних із самою комунікацією (Анонім, 2018?, с.7). Це не просто механічне вживання слів, а вміле й гнучке керування ходом комунікації, що дозволяє мовцю ефективно впливати на співрозмовника. Стратегія мовленнєвого спілкування базується на розумінні конкретної ситуації спілкування, намірів мовця та очікувань співрозмовника. Оратор досягне бажаного ефекту комунікації, якщо ретельно спланує свою мовленнєву поведінку, спрогнозує її успіх (тобто визначить стратегію) і вибере поетапні дії (тактичні прийоми) для досягнення своєї мети (Авраменко, 2018, с.68).

У сучасній лінгвістиці не існує єдиного підходу до визначення стратегій і тактик мовної маніпуляції. Автор праці «Стратегії мовного маніпулювання у виступах політичних лідерів» зазначає, що деякі українські вчені, такі як В.В.

Зірка та О.В. Дмитрук та інші вважають терміни «стратегія» й «тактика» тотожними. На противагу їм, дослідники В.В. Одинцов, Н.І. Формановська та інші, акцентують увагу на відмінності значень цих термінів (Анонім, 2018?).

О.А. Новицька (2018, с.23-24) розділяє та визначає ці поняття так: **комунікативна стратегія** – це сукупність запланованих мовцем і реалізованих в ході комунікативного акту теоретичних ходів, спрямованих на досягнення комунікативної мети. Тоді як **комунікативна тактика** – це сукупність практичних ходів у реальному процесі мовленнєвої взаємодії.

О.Л. Ляпічева (2016) пропонує такі тлумачення: **стратегія мовного спілкування** – це оптимальна реалізація намірів мовця щодо досягнення конкретної мети; контроль і вибір ходів спілкування, їх гнучкої видозміни в конкретній ситуації. **Тактика спілкування** – це певна лінія поведінки на конкретному етапі комунікативної взаємодії; мовні прийоми досягнення мети.

Залежно від поставлених завдань і контексту взаємодії, дослідники пропонують різні способи класифікації комунікативних стратегій. Так, Т. Толмачова (Маслова, н. д., с.83) розробляє типологію стратегій, що базується на мовленнєвих актах, згрупованих за їх **комунікативною функцією** – *передача інформації, коментування, емоційне вираження, спонукання до дій, регуляція стосунків між співрозмовниками та підтримка уваги аудиторії*. Відповідно, **інформативна стратегія** стосується *повідомлення або отримання інформації*, іноді з прихованим впливом на поведінку адресата; **оцінно-впливова** передбачає *вербальне вираження оцінок і суджень* для формування певного ставлення; **емоційно-впливова** полягає у *вираженні емоцій*, щоб змінити *психоемоційний стан співрозмовника*; **регулятивно-спонукальна** спрямована на *спонукання до дії через поради, вимоги чи накази*; а **конвенційна** охоплює *ритуальні етикетні дії й підтримання контакту між учасниками спілкування*.

Іншого підходу дотримується Л. Д. Швелідзе (2021), яка аналізує комунікативні стратегії на матеріалі соціальних мереж. Авторка виокремлює **кооперативні стратегії аргументації** (реалізується за допомогою тактик *авторитетності, діалогічності, суб'єктивності та солідаризації*),

інформаційності (описова, констатувальна та фактографічна тактики), **самопрезентації** (самооцінювання, гіперболізація та ідеалізація), **спонукання** (основні тактики – наказ, прохання і рекомендація) та **ритуальності** (привітання, подяку та співчуття), а також **конфліктні стратегії дискредитації** (звинувачення, образи та погрози), **незгоди** (представлена в соцмережах тактиками критики і самокритики) і **тролінгу** (спирається на мовленнєву провокацію та конфронтацію).

Залежно від **мети** існують різні **мовленнєві стратегії**: мовець намагається *завоювати довіру когось, дискредитувати, заперечити відомі факти, спонукати до схвалення чи неприйняття, закликати до дії, співпраці, ворожнечі чи дружби*. Відповідно до цього здійснюється вибір **тактики мовної поведінки**: *лестоці, провокації, залякування, внесення неформальностей, гумор, маніпулювання фактами та емоціями* (Авраменко, 2018, с.69).

За своєю **функціональністю** виділяють **основні й допоміжні** стратегії. Основні стратегії є найважливішими для комунікативних взаємодій з точки зору мотивів та цілей. Допоміжні стратегії сприяють ефективній організації мовленнєвої взаємодії, наприклад, стратегії статусу й ролі, спрямовані на створення емоційного настрою тощо (Новицька, 2018, с.24).

Мовні тактики функціонують як засоби реалізації мовної стратегії: за їх допомогою формуються окремі фрагменти діалогу, в яких поєднуються та змінюються модальні компоненти висловлення – оцінки, судження, припущення тощо (Андрущенко, Скубашевська, 2010, с.138).

Отже, поняття мовних стратегій охоплює широкий діапазон підходів, що відображають складність і багатовимірність комунікативного процесу. У науковій літературі немає єдиного тлумачення цього явища, проте загальноприйнято розглядати стратегію як свідомо сплановану послідовність мовленнєвих дій, спрямовану на досягнення конкретної мети. Різні класифікації, засновані на функціональних, ситуаційних чи жанрових критеріях, підкреслюють динамічність мовних стратегій і необхідність їх гнучкого застосування в різних умовах спілкування.

1.2. Особливості цифрового медіапростору та Telegram як платформа реалізації стратегій впливу

Сучасне інформаційне суспільство характеризується стрімким зростанням потенціалу засобів масової інформації (медіа) та посиленням їх впливу на свідомість і підсвідомість людини. У цифровому просторі мовні стратегії набувають особливого значення як засіб ефективного впливу на масову аудиторію.

У науковій літературі часто використовують два терміни: «**цифрове середовище**» і «**мережа Інтернет**». Тому важливо зрозуміти їх взаємозв'язок. Цифрове середовище є більш загальним поняттям порівняно з мережею Інтернет, оскільки воно включає не лише вебсайти та вебсторінки, але й електронні документи, файли, а також інші об'єкти, які використовуються на різних пристроях (комп'ютерах, ноутбуках, планшетах, телефонах, гаджетах), що не мають паперових аналогів (Тарасенко, 2019).

Мовні стратегії застосовуються для досягнення різноманітних цілей – від інформування до формування певних переконань та установок. У цифровому середовищі (соціальні мережі, месенджери, відеоплатформи) ці процеси посилюються через швидкість поширення інформації, емоційність подачі та можливість таргетованого впливу на конкретну аудиторію.

Посмітна В. В. (2019, с.124) зазначає **загальні комунікативні цілі масмедіа**. По-перше, **формування громадської думки та поширення певних ідей у суспільстві**, що, по суті, визначає напрям подальшої діяльності медіа і впливає на вибір комунікативних стратегій. По-друге, **висвітлення важливої інформації**, мета, яка лежить на поверхні, і звичайні громадяни очікують саме актуальної інформації. Водночас представники ЗМІ самостійно вирішують, яка інформація є важливою в даний момент. По-третє, **контролювання влади, державних структур, захист прав громадян** – такі цілі діяльності ставлять перед собою самі журналісти, позиціонуючи ЗМІ як четверту владу між суспільством і керівництвом держави.

В цій праці також було схарактеризовано **комунікативні можливості масмедіа**. **Можливість швидкого поширення інформації** є однією з визначальних рис дискурсу засобів масової інформації. З одного боку, це значною мірою сприяє ефективності впливу, а з іншого – вимагає необхідності постійного оновлення інформації та її якнайшвидшої публікації, що створює загрозу поширення не підтверджених і недостовірних даних.

Здатність висвітлювати суб'єктивні думки відрізняє дискурс ЗМІ від дискурсу наукового та офіційно-ділового, що становить головну загрозу у формуванні упередженої оцінки реальних явищ.

Крім того, засоби масової інформації можуть **впливати на емоційну сферу**. Як відомо, якщо вплив здійснюється як на логічному, так і на емоційному рівні, то він має значно вищу ефективність, ніж вплив був направлений не лише на розумову діяльність індивіда (Посмітна, 2019, с.123).

У статті «Мовні засоби впливу на читача у сучасних ЗМІ» (Матвійчук, Житар, 2024) автори наголошують на потужній **комунікативно-впливовій функції медіа**. Для реалізації цієї функції застосовуються різноманітні мовні стратегії та риторичні прийоми. Серед них – *використання емоційно забарвленої лексики, сенсаційних заголовків, узагальнень, двозначних формулювань*, що створюють емоційний або когнітивний тиск на читача. Часто інформація подається вибірково, із *підкресленням одних фактів та замовчуванням інших*, що дозволяє маніпулювати сприйняттям реальності. Поширеними засобами такого впливу є *кліше, сленгові вирази, емоційно-експресивна лексика, лозунгові формулювання, стилістичні фігури, фразеологізми, інішомовні запозичення*, а також активне використання *пунктуаційних знаків для акцентування*. Крім цього, для журналістів важливо влучно підібрати мовні засоби, враховуючи *тематику та цільову аудиторію*, що підвищує ефективність комунікації.

Також значну роль у цифровому середовищі відіграють **алгоритми платформ**, які не лише відображають контент, а й активно формують інформаційне поле користувача. Рекомендаційні системи TikTok, YouTube чи

Instagram підсилюють вплив, підбираючи контент, що відповідає попереднім вподобанням, – це створює ефект «інформаційної бульбашки», у якій користувач рідко стикається з альтернативними поглядами. Таким чином, вплив медіа посилюється не тільки через сам контент, а й через спосіб його подачі, який контролює платформа.

Як зазначає голова групи з аналізу гібридних загроз Українського кризового медіацентру, Олександра Цехановська, **інформаційна бульбашка** – це стан інформаційної ізоляції. Часто люди опиняються у ньому добровільно, отримуючи ту інформацію, яка їм подобається і з якою вони вже згодні. На її думку, причиною цього явища є не лише алгоритми платформ, а й природні когнітивні упередження людини, які підштовхують мозок сприймати інформацію, що не суперечить власній картині світу (Громадське радіо, 2022).

Це підтверджує, що мовний вплив у цифровому середовищі не функціонує ізольовано. Він тісно пов'язаний та існує одночасно з технічними механізмами поширення інформації – алгоритмами, які формують інформаційне поле та підсилюють ефект від стратегії впливу. Це створює умови для глибокого впливу на свідомість користувача, підсилюючи ефективність використаних мовних засобів.

Особливе місце в сучасному цифровому медіапросторі України посідає платформа Telegram (Telegram, n. d.-b). Данько-Сліпцова А. (2023) зазначає, що повномасштабне вторгнення 2022 року кардинально змінило інформаційні звички українців: попри значну цифровізацію суспільства та зусилля найбільших телеканалів, які об'єдналися в цілодобовий телемарафон, традиційні ЗМІ не встигали задовольняти запит на миттєву інформацію. Саме тоді на перший план вийшли Telegram-канали. Серед їхніх ключових переваг дослідниця виокремлює легкість отримання та поширення інформації, можливість завантаження навіть при мінімальному інтернет-з'єднанні, відсутність цензури та можливість створювати канали з необмеженою кількістю підписників (Данько-Сліпцова, 2023, с. 275).

Масштаб цього зсуву підтверджують соціологічні дані. За результатами опитування Київського міжнародного інституту соціології, проведеного у липні 2022 року, соціальні мережі вважають головними джерелами новин 59% респондентів, а майже половина всього часу на споживання новин (48%) припадає саме на соціальні мережі. Серед них лідером став Telegram – на нього витрачається 41% часу користувачів у соціальних мережах (Данько-Сліпцова, 2023, с. 280).

Водночас дослідниця наголошує на принциповій відмінності між Telegram-каналами та повноцінними медіа: канали не є зареєстрованими засобами масової інформації, не несуть юридичної відповідальності за достовірність контенту і не зобов'язані дотримуватися журналістських стандартів (Данько-Сліпцова, 2023, с. 279). Це створює умови для безконтрольного поширення непідтвердженої інформації та маніпулятивних повідомлень, що робить вивчення вербальних стратегій впливу саме в цьому середовищі особливо актуальним.

Отже, цифровий медіапростір є потужним середовищем для реалізації вербальних стратегій впливу, оскільки поєднує мовні засоби з технічними механізмами поширення інформації. Його гнучкість, оперативність та широке охоплення аудиторії сприяють формуванню суспільної думки, впливаючи одночасно на когнітивному та емоційному рівнях. Алгоритми платформ, що формують індивідуальне інформаційне поле користувача, додатково підсилюють ефект застосованих стратегій, створюючи умови для глибокого впливу на свідомість. В українських реаліях воєнного часу ці процеси набули особливої гостроти: Telegram як домінуюча новинна платформа поєднує всі зазначені властивості цифрового середовища з відсутністю редакційного контролю та юридичної відповідальності авторів, що створює особливо сприятливі умови для цілеспрямованого застосування вербальних стратегій впливу. Це й зумовлює актуальність їх систематичного вивчення та розробки інструментів автоматизованого виявлення.

1.3. Маніпуляція в медіа: техніки, цілі та адаптація до аудиторії

Важливим аспектом у дослідженні комунікаційних стратегій медіа є розрізнення **деструктивного і недеструктивного мовного впливу**. Такі впливи розрізняють за моральним критерієм. Мовний вплив, створений з метою мати шкідливі наслідки для адресата, вважаємо **деструктивним**. При виявленні засобів комунікації деструктивного впливу в засобах масової інформації важливо враховувати основні цілі їх діяльності, їх комунікативні здібності, а також комунікативні стратегії і тактику медійного дискурсу (Посмітна, 2019, с.123).

У контексті інформаційної війни науковці приділяють особливу увагу деструктивному впливу масмедіа. Безвідповідальна журналістика може відігравати ключову роль у розпалюванні ворожнечі в суспільстві, а неперевірена інформація чи некоректне формулювання можуть стати об'єктом маніпулювання ворожих ЗМІ (Анонім, 2005).

Серед уже розглянутих в пункті 1.2. комунікативних стратегій масмедійного дискурсу важливо детальніше розглянути роль **емоційної та маніпулятивної стратегії**, які істотно впливають на свідомість та поведінку аудиторії.

Для кращого розуміння цих стратегій зазначимо, що **маніпулювання** – це вплив на особу з метою спонукати її до вчинків несвідомо або наперекір її власній думці, задуму. На відміну від прямого примусу, який є відкритим і очевидним, маніпулятивний вплив має прихований (імпліцитний) характер, тож адресат не завжди розпізнає сам факт впливу (Новицька, 2018, с.24-25).

Стратегія емоційного впливу передбачає цілеспрямоване використання мовних і позамовних засобів з метою формування стійких стереотипів, асоціацій і переконань, а також коригування світоглядних орієнтирів. У цьому контексті ЗМІ нерідко **спекулюють на сильних емоціях**, таких як страх, відповідальність, патріотизм – вдаючись до тактик *залякування, шокуючих прогнозів, психологічного тиску чи емоційного зараження*. Такі підходи можуть мати **небезпечні наслідки** як на **індивідуальному рівні** (*зниження критичного*

мислення, емоційна нестабільність, агресивна поведінка), так і на **суспільному** (загострення конфліктів, суспільна напруга, поширення деструктивних наративів).

Не менш загрозливою в межах інформаційного простору є **стратегія маніпулювання**, яка часто поєднується з іншими формами впливу. Її основною ознакою є прихований характер – як у **викладенні інформації** (через вибірковість, перекручення чи замовчування фактів), так і в **передачі намірів комунікатора**. Такий прихований вплив ускладнює ідентифікацію маніпуляції, що підсилює її ефективність і водночас небезпеку (Посмітна, 2019, с.125-126).

Таке використання мовних засобів, що спрямовані на раціональні (логіка, розум) та ірраціональні (почуття, віра, підсвідомість) «мішені», забезпечує ефективну реалізацію маніпулятивної стратегії. Такий вплив є послідовним процесом, що включає стратегічне планування, вибір тактик і засобів, запуск механізмів впливу та досягнення перлюкутивного ефекту.

Значну роль у цьому відіграють **заголовки**, які є першою точкою контакту з інформацією. Вони можуть не лише інформувати, а й впливати на сприйняття події, емоційно налаштовувати читача або ж спрямовувати увагу на певні аспекти, залишаючи інші поза фокусом. Часто заголовки формуються із застосуванням *експресивної лексики, кліше, іншомовних запозичень, неологізмів та слів у переносному значенні*, що посилює їхню маніпулятивну функцію. Таким чином, медійний дискурс активно використовує мовні стратегії та засоби, спрямовані на формування уявлень, емоцій і поведінкових реакцій аудиторії, що свідчить про його здатність впливати не лише на інформованість, а й на світогляд читача (Матвійчук, Житар, 2024, с. 40).

З позиції соціальних наук маніпуляцію розглядають як сукупність засобів ідеологічного та соціально-політичного впливу, спрямованих на трансформацію свідомості та поведінки людей у напрямку, що суперечить їхнім власним інтересам. При цьому індивіди часто не усвідомлюють, що їхні погляди, потреби, пріоритети й стиль життя значною мірою формуються під впливом зовнішніх сил, які здійснюють маніпуляцію.

На основі цього підходу Любимий Я. В. (1993) виокремлює два головні **типи маніпулятивного впливу**: оперативний та стратегічний.

Оперативна (або ситуативна) маніпуляція використовує вже наявні у свідомості індивідів установки – цінності, потреби, стереотипи, звички – для того, щоб викривити сприйняття інформації у вигідний для маніпулятора спосіб. Внаслідок цього широкі верстви населення починають діяти відповідно до запланованого сценарію: *беруть участь або ігнорують події, виходять на протести, створюють хаос або, навпаки, демонструють злагожденість та підтримують порядок.*

Стратегічна маніпуляція, у свою чергу, передбачає *довготривалий вплив на свідомість*, у процесі якого поступово закріплюються вигідні для маніпулятора цінності, ідеї, моделі поведінки й мислення. У перспективі саме ці закріплені установки можуть стати підґрунтям для оперативної маніпуляції, коли виникне така необхідність.

У сучасних медіа особливо помітною є тенденція до **фрагментації подачі інформації**: матеріали часто подаються не послідовно, а у вигляді коротких, емоційно насичених фрагментів. Це створює ілюзію різноманіття й оперативності, хоча насправді може викликати спотворене уявлення про реальність. При цьому активно застосовуються прийоми впливу на підсвідомість – від тонких лінгвістичних маніпуляцій до створення візуальних образів із заданим емоційним відтінком. **Маніпулятивність** може виявлятися через *викривлення масштабів подій, подачу їх у сенсаційній формі, використання символічного ряду або апеляцію до колективних емоцій*, що робить інформаційний вплив значно ефективнішим і водночас менш помітним для реципієнта (Анонім, 2011).

Після розгляду маніпулятивних стратегій у медіа логічним наступним кроком є аналіз того, як ці стратегії адаптуються до різних аудиторій. Ефективність медіакомунікації значною мірою залежить від здатності розуміти та враховувати специфіку цільових груп, що досягається через сегментацію аудиторії.

Сегментація аудиторії медіа – це поділ загальної аудиторії на менші групи за спільними ознаками: *віком, статтю, освітою, доходом, стилем життя та інтересами*. Такий підхід дозволяє адаптувати контент до потреб конкретних соціальних груп, роблячи його більш релевантним і впливовим. Наприклад, молодь активніше сприймає теми цифрових технологій і соціальних мереж, тоді як старші люди більше цікавляться питаннями здоров'я та добробуту. Водночас стать також впливає на вподобання споживачів медіа – чоловіки та жінки нерідко мають різні інформаційні запити. Доходи й освітній рівень аудиторії зумовлюють стиль викладу, глибину аналітики та тематику матеріалів. Крім того, сучасні медіа враховують стиль життя аудиторії: контент може орієнтуватися на екосвідомих, техноорієнтованих чи, наприклад, спортивно активних людей. Усе це сприяє підвищенню залученості й довіри до медійного продукту.

Сегментація аудиторії має вирішальне значення для сучасних новинних медіа, оскільки дозволяє створювати контент, який максимально відповідає інтересам різних груп споживачів інформації. Завдяки чіткому розумінню соціально-демографічних та поведінкових характеристик аудиторії, журналісти й редактори можуть адаптувати новинні повідомлення так, щоб вони були релевантними для конкретних цільових сегментів. Наприклад, молодіжна аудиторія очікує стислих, динамічних новин у форматі відео чи сторіс, тоді як старші групи можуть надавати перевагу розгорнутим текстам із глибшим аналітичним змістом. Такий підхід не лише сприяє зростанню залученості та довіри до джерела, а й дає змогу ефективніше використовувати ресурси редакції, спрямовуючи зусилля на створення саме того контенту, який резонує з потребами кожного сегмента. Крім того, систематичне спостереження за змінами у вподобаннях аудиторії дає змогу медіа оперативно реагувати на нові тренди й залишатися актуальними в конкурентному середовищі (Unknown, 2023).

Таким чином, маніпулятивні стратегії у медіа-комунікації становлять серйозну загрозу в інформаційному просторі, особливо в умовах соціальної

напруги та війни. Їхня ефективність полягає у прихованому характері впливу, що поєднує раціональні та емоційні чинники, мовні й візуальні прийоми. Особливу небезпеку становлять стратегії, орієнтовані на масову аудиторію, яка часто не усвідомлює маніпулятивного характеру повідомлення. Водночас адаптація контенту до специфіки цільових груп лише посилює вплив – завдяки точному врахуванню потреб, цінностей і стилю життя кожного сегмента.

Висновки до першого розділу

Отже, перший розділ бакалаврської роботи розкрив теоретичні основи мовних стратегій у медіа, визначивши їх як свідомо сплановані дії для досягнення комунікативних цілей. Аналіз показав, що немає єдиного підходу до класифікації стратегій, але їх поділяють за функціями (інформативна, емоційно-впливова тощо), ситуаціями чи жанрами, що підкреслює їхню гнучкість. Особливу увагу приділено цифровому медіапростору, де Telegram завдяки оперативності поширення інформації, анонімності авторів і відсутності редакційного контролю створює особливо сприятливі умови для реалізації стратегій впливу. Алгоритмічне формування інформаційного поля та емоційна подача матеріалу додатково підсилюють ефект застосованих стратегій, сприяючи формуванню громадської думки. Також розглянуто деструктивний вплив у медіа – зокрема маніпуляцію та емоційний тиск, – які набувають особливої гостроти в умовах війни та інформаційного протистояння. Таким чином, розділ заклав теоретичну базу, що демонструє, як цифрове середовище трансформує вербальні стратегії, роблячи їх ефективнішими, але й потенційно небезпечними. Ці положення стануть основою для практичного аналізу із застосуванням методів машинного навчання в наступних розділах роботи.

РОЗДІЛ 2. ФОРМУВАННЯ КОРПУСУ ДАНИХ ТА МЕТОДОЛОГІЯ ДОСЛІДЖЕННЯ

У цій частині бакалаврської роботи на основі теоретичних засад, викладених у першому розділі, сформовано емпіричну базу дослідження: визначено та обґрунтовано вибір джерел, описано процес збору й структурування даних, а також здійснено ручну розмітку корпусу повідомлень відповідно до виокремленої класифікації вербальних стратегій впливу.

2.1. Методи дослідження та огляд матеріалів для аналізу

Дослідження буде базуватися на методах контент- та дискурс-аналізу.

Контент-аналіз – це техніка для якісного та кількісного аналізу вмісту документів з метою виявлення або вимірювання різних фактів і тенденцій, зафіксованих у документах. Особливістю контент-аналізу є дослідження документів відповідно до їх соціального контексту (Лук'янова, 2010, с.1).

Дискурс-аналіз – це набір методів і прийомів інтерпретації різних типів текстів як продуктів мовної діяльності, що здійснюється в конкретних культурно-історичних умовах. Т. А. Ван Дейк писав: «В широкому сенсі дискурс є складною єдністю мовленнєвої форми, значення і дії, яка найкраще може бути охарактеризована за допомогою поняття комунікативної події (акту)» (Комісар, с. 1). Тобто з його допомогою розкривається прихований зміст тексту, підґрунтя його створення, можливі інтерпретації читача/слухача тощо (Семигіна, 2001, с.1).

Хоча контент-аналіз і дискурс-аналіз досліджують соціальну реальність, обидва методи базуються на різних уявленнях про природу цієї реальності загалом і про роль мови зокрема. Якщо **дискурс-аналіз** зосереджується на *мінливій і непостійній природі* значення та його дослідженні, **контент-аналіз** виходить з *постійності й незмінності* значення. Це дозволяє вважати слова та тексти повністю еквівалентними реальності та дає можливість їх кількісного інтерпретаційного аналізу. Дискурс-аналіз фокусується на взаємозв'язку між

текстом і контекстом, тоді як контент-аналіз інтерпретує текст абстрактно від контексту (Семотюк, 2015, с. 2).

Виходячи з теоретичної бази та обраних методів дослідження, для аналізу новинного контенту було обрано платформу Telegram як основне середовище дослідження (Telegram, n. d.-b). Як було обґрунтовано в підрозділі 1.2, саме ця платформа стала домінуючим джерелом новин для українців в умовах повномасштабної війни, поєднуючи оперативність поширення інформації з відсутністю редакційного контролю – що створює особливо сприятливі умови для реалізації різноманітних вербальних стратегій впливу.

Для формування вибірки було проаналізовано рейтинг найпопулярніших українських Telegram-каналів за даними порталу galera.news (Крамар Сара, 2025). З цього переліку було відібрано чотири канали, а п'ятий – авторського типу – було взято для типологічного порівняння. Таким чином, вибірка охоплює три типи каналів: анонімні агрегатори новин, офіційний державний канал та авторський канал з верифікованим автором. За даними сайту galera.news:

- Труха ⚡ Україна (Труха Україна, н. д.) – один із найпопулярніших українських Telegram-каналів, що розпочинав як локальний харківський ресурс, а згодом виріс у загальнонаціональний інформаційний майданчик. Канал оперативно висвітлює резонансні події, надзвичайні ситуації, політичні рішення та бойові дії.
- Реальна Війна | Україна | Новини (Реальна Війна | Україна | Новини, н. д.) – канал, що спеціалізується на висвітленні бойових дій: публікує щоденні зведення, карти фронту, фото- та відеоматеріали, документуючи перебіг війни в режимі реального часу.
- Телеграмна служба новин - Україна (Телеграмна служба новин - Україна, н. д.) – канал широкого тематичного охоплення, що інформує читача про події на фронті, в економіці, політиці та культурі у форматі стислих повідомлень.
- Zelenskiy / Official (Zelenskiy / Official, н. д.) – офіційний канал президента України Володимира Зеленського, на якому першими

з'являються його звернення, промови та повідомлення про офіційні заходи. Канал є виразником державного наративу і слугує прямим комунікаційним містком між владою та суспільством.

- Лачен пише (Лачен пише, н. д.) – авторський канал журналіста Ігоря Лаченкова, присвячений українській, російській та міжнародній політиці. На відміну від решти каналів вибірки, він веде відкриту авторську позицію та публікує матеріали від власного імені, що робить його цінним об'єктом для порівняльного аналізу вербальних стратегій (Федотова, 2024).

Така типологічна різноманітність вибірки дозволяє порівняти вербальні стратегії впливу, характерні для різних суб'єктів українського медіапростору, та виявити закономірності, що не були б помітні при аналізі однотипних джерел.

Період мого дослідження охоплює період грудень 2023 – січень 2024 року. Цей вибір не є випадковим і має кілька підстав. З одного боку, він є достатньо віддаленим від початку повномасштабного вторгнення, щоб більшість українських Telegram-каналів уже завершили повний перехід на українську мову публікацій. Саме ця умова є важливою для проведення коректного лінгвістичного аналізу. З іншого боку, цей період залишається достатньо близьким до початку війни, і саме з цієї причини медіапростір все ще має специфічну інтенсивність інформаційного тиску воєнного часу: високий рівень емоційності у викладі матеріалів, активне використання маніпулятивних прийомів та риторика мобілізації. Таким чином, обраний період поєднує мовну однорідність корпусу з насиченістю вербальними стратегіями впливу, що робить його оптимальним для цілей цього дослідження.

Отже, методологічну основу дослідження утворюють два взаємодоповнювальні методи – контент-аналіз та дискурс-аналіз. Контент-аналіз надає кількісне вимірювання повторюваних мовних тенденцій, тоді як дискурс-аналіз дає змогу розкрити прихований зміст повідомлень та досягнути їх у соціально-культурному тлі воєнного часу. Дослідницьку базу становлять матеріали п'яти українських Telegram-каналів, обраних за критерієм

типологічної різноспрямованості: анонімні агрегатори новин, офіційний державний канал та авторський канал із підтвердженням автором. Така добірка уможлиблює зіставлення вербальних стратегій різних суб'єктів медіапростору та виявлення закономірностей, непомітних при дослідженні однорідних джерел. Часові межі – грудень 2023 – січень 2024 року – визначені як оптимальні, бо поєднують мовну сталість корпусу зі збереженням притаманної воєнному часу гостроти інформаційного впливу, що формує стабільну основу для подальшого розгляду.

2.2. Програмна реалізація системи автоматизованого збору та попередньої обробки даних

Для проведення лінгвістичного дослідження стратегій у медіадискурсі було розроблено програму для автоматизованого збору й початкової обробки текстових даних із Telegram-каналів. Результатом роботи на цьому етапі є структурована база даних, яка містить корпус текстів, придатний для подальшої ручної та автоматизованої розмітки. Весь програмний код розміщено у відповідному GitHub репозиторії (див. Додаток 1). Нижче описано архітектуру скрипту та логіку його окремих компонентів.

- **Використані бібліотеки та їхнє призначення**

Код написано в середовищі PyCharm (JetBrains, n. d.) мовою програмування Python (Python Software Foundation, n. d.). Для реалізації програми було використано декілька бібліотек (рис. 2.2.1).

```
from telethon import TelegramClient
from datetime import datetime
from zoneinfo import ZoneInfo
import sqlite3
import asyncio
import re
from tqdm import tqdm
```

Рисунок 2.2.1. Імпортування потрібних бібліотек

Центральною з них є *Telethon* (Lonami, n. d.) – асинхронна бібліотека для роботи з офіційним Telegram API. На відміну від стандартних бот-інтерфейсів,

вона працює через протокол MTProto (Telegram, n. d.-a) та надає доступ до повного функціоналу клієнтської частини Telegram: читання повідомлень будь-яких публічних каналів, отримання метаданих, ітерація по історії повідомлень. Саме ця можливість робить її корисною для задач корпусної лінгвістики та контент-аналізу.

Модулі *datetime* і *zoneinfo* відповідають за коректну обробку часових міток повідомлень з урахуванням часового поясу.

Бібліотека *sqlite3* є вбудованим модулем Python і забезпечує підключення до реляційної бази даних SQLite (SQLite Consortium, n. d.) – легкої файлової СУБД, що не потребує окремого серверного процесу. Уся база даних зберігається в одному файлі *telegram_messages.db*, що спрощує її перенесення та подальше використання в аналітичних інструментах.

Модуль *asyncio* реалізує механізм асинхронного програмування в Python, що дозволяє запускати кілька задач одночасно в межах одного потоку. У даному контексті це означає, що збір повідомлень із п'яти каналів відбувається паралельно, а не послідовно, що суттєво скорочує загальний час виконання програми.

Бібліотека *re* (регулярні вирази) забезпечує лінгвістичне очищення текстів: видалення рекламних вставок, гіперпосилань, Markdown-розмітки та зайвих символів.

Бібліотека *tqdm* забезпечує виведення інтерактивного індикатора прогресу в терміналі під час збору даних, що дозволяє відстежувати поточний стан виконання програми.

- **Налаштування доступу до Telegram API**

Для підключення до Telegram API необхідно зареєструвати власний застосунок на платформі my.telegram.org, після чого отримати два унікальні ідентифікатори: *api_id* та *api_hash* (рис. 2.2.2).

App configuration

App api_id:	*****	🔒
App api_hash:	*****	🔒
App title:	Testing App 1	
Short name:	testapp123	
	alphanumeric, 5-32 characters	

Рисунок 2.2.2. Реєстрація на платформі my.telegram.org

Ці параметри передаються при ініціалізації об'єкта *TelegramClient*, який є точкою входу для всіх запитів до платформи. Перший запуск програми ініціює сеанс авторизації (session) (рис. 2.2.3), який зберігається локально у файлі `session.session` і надалі використовується повторно без додаткового підтвердження.

```
Підключення до БД...
БД готова

Початок збору даних...
Please enter your phone (or bot token): +380
Please enter the code you received: 53495
```

Рисунок 2.2.3. Процес первинної авторизації

- **Формування дослідницької вибірки**

Оскільки об'єктом аналізу є україномовне середовище у період з грудня 2023 по січень 2024 року, у коді реалізовано строгу фільтрацію за часом за допомогою модуля *datetime*. Особлива увага приділена коректній роботі з часовими поясами: використання бібліотеки *zoneinfo* дозволяє нормалізувати час публікації повідомлень відповідно до київського часового поясу (Europe/Kyiv), що дає змогу встановити чіткі рамки для зібраного матеріалу (рис. 2.2.4).

```
# Встановлення часового поясу та хронологічних меж дослідження
ukraine_tz = ZoneInfo("Europe/Kyiv")
start_date = datetime( year: 2023, month: 12, day: 1, hour: 0, minute: 0, second: 0, tzinfo=ukraine_tz)
end_date = datetime( year: 2024, month: 1, day: 31, hour: 23, minute: 59, second: 59, tzinfo=ukraine_tz)
```

Рисунок 2.2.4. Встановлення часових рамок через модуль *datetime*

- **Функція лінгвістичного очищення тексту**

Функція *clean_text()* реалізує кілька послідовних етапів нормалізації тексту. Її головна мета – перетворити «сире» повідомлення з Telegram на чистий текстовий матеріал, придатний для машинного навчання. На першому етапі за допомогою регулярних виразів вилучаються типові рекламні вставки, посилання на інші ресурси та заклики до підписки, характерні для певних каналів. Далі видаляються гіперпосилання у форматі Markdown та прямі URL-адреси, оскільки вони не несуть лінгвістичної цінності для аналізу. Після цього текст очищується від нелітерних символів із збереженням базової пунктуації та нормалізуються зайві пробіли (рис. 2.2.5).

```
# Функція для лінгвістичного очищення та первинної нормалізації тексту
def clean_text(text): 1 usage
    if not text:
        return None

    text = re.sub(
        pattern: r'(?i)(ТРУХА.*?Украина.*?Прислать новость|ТРУХА.*?Україна.*?Надіслати новину|Реальна Війна
        repl: '', text, flags=re.DOTALL)

    # Markdown-посилання → залишаємо текст
    text = re.sub( pattern: r'\[[^\]]+\]\(https?:/[^\]]+\)', repl: r'\1', text)

    # Тільки явні URL зі схемою – безпечно і достатньо
    text = re.sub( pattern: r'https?:/\S+', repl: '', text)
    # t.me окремо – без схеми, але точний патерн
    text = re.sub( pattern: r't.me/\S+', repl: '', text)

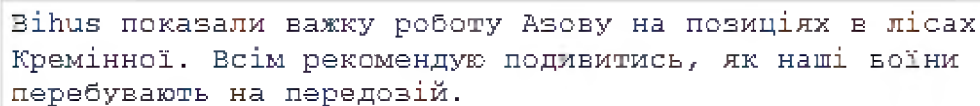
    text = re.sub( pattern: r'[\w\s.,!?\'"(){}[\]\<>:\;\---«»/\\"$%&&+##~°]', repl: '', text)
    text = re.sub( pattern: r'[_\[\]]', repl: '', text)
    text = re.sub( pattern: r'\s+', repl: ' ', text).strip()

    if not text:
        return None
```

Рисунок 2.2.5. Реалізація алгоритму лінгвістичного очищення

Окремо реалізовано мовний фільтр, який працює у кілька етапів. Спочатку перевіряється наявність кириличних символів у тексті: якщо їх немає взагалі –

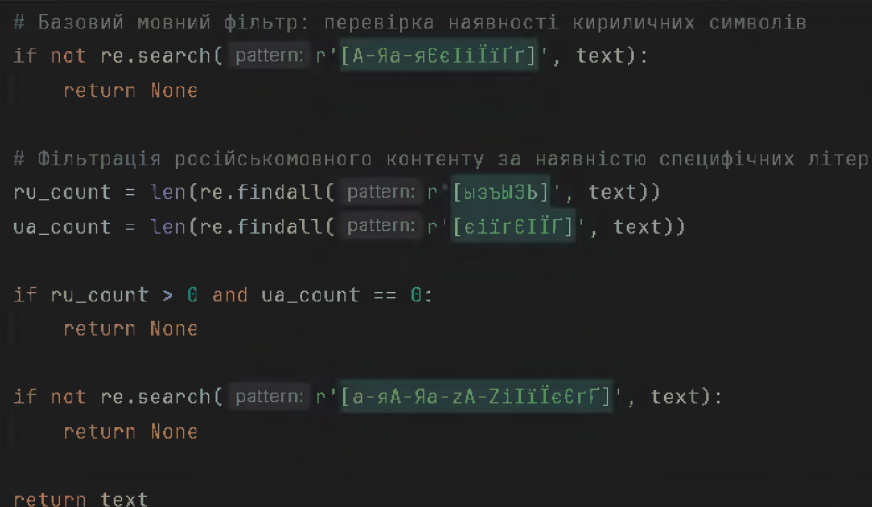
повідомлення відхиляється. Це автоматично усуває з корпусу тексти, написані виключно латиницею, зокрема повністю англomовні дописи на каналі Президента (Zelenskiy / Official). Водночас повідомлення, що містять україномовний текст із вкрапленнями іншомовних слів, термінів чи аббревіатур, успішно проходять цей фільтр, оскільки кирилиця в них присутня (рис. 2.2.6). Таким чином зберігається природна мовна специфіка україномовного медіадискурсу, де іншомовні вставки є нормою, а не виключенням.



`Vihus` показали важку роботу Азову на позиціях в лісах
Кремінної. Всім рекомендую подивитись, як наші воїни
перебувають на передовій.

Рисунок 2.2.6. Приклад повідомлення з латинськими вкрапленнями після мовного фільтру

На другому етапі здійснюється розмежування україно- та російськомовного контенту. Для цього підраховується кількість літер, специфічних для кожної з мов: символи ы, э, ь є маркерами російської мови, тоді як є, і, ї, г – виключно українськими. Якщо повідомлення містить лише російськомовні маркери без жодного україномовного, воно не додається до корпусу. Завершальна перевірка додатково підтверджує наявність у тексті хоча б одного літерного символу – кириличного або латинського, – що виключає записи, які складаються лише з цифр, пунктуації, спецсимволів чи емодзі. Таким чином забезпечується мовна однорідність дослідницької вибірки (рис. 2.2.7).



```
# Базовий мовний фільтр: перевірка наявності кирилических символів
if not re.search( pattern: r'[А-Яа-яЄєІіЇїґ]', text):
    return None

# Фільтрація російськомовного контенту за наявністю специфічних літер
ru_count = len(re.findall( pattern: r'[ыэьЭэ]', text))
ua_count = len(re.findall( pattern: r'[єіїґЄІЇґ]', text))

if ru_count > 0 and ua_count == 0:
    return None

if not re.search( pattern: r'[а-яА-Яа-зА-ЗііїґЄєґ]', text):
    return None

return text
```

Рисунок 2.2.7. Реалізація алгоритму мовного відбору

- **Асинхронний збір повідомлень**

Функція *fetch_channel()* є асинхронною (позначена ключовим словом *async*) і виконує збір повідомлень із одного каналу. Спочатку отримується офіційна назва каналу через метод *get_entity()*, яка надалі зберігається у полі **Джерело** замість URL-адреси. Ітерація по повідомленнях відбувається у зворотному хронологічному порядку – від найновіших до найдавніших – за допомогою методу *iter_messages()* з параметром *offset_date*, що задає верхню часову межу. Як тільки дата чергового повідомлення стає меншою за встановлену нижню межу, цикл переривається (рис. 2.2.8).

```
# Асинхронна функція для вивантаження повідомлень із зазначеного каналу
async def fetch_channel(client, channel_name, results, pbar): 1 usage
    try:
        # Отримання метаданих каналу (офіційної назви)
        entity = await client.get_entity(channel_name)
        channel_title = entity.title
        count = 0
        # Ітеративний збір повідомлень у зворотному хронологічному порядку
        async for message in client.iter_messages(entity, offset_date=end_date):
            # Припинення збору при досягненні нижньої хронологічної межі
            if message.date < start_date:
                break
```

Рисунок 2.2.8. Фрагмент функції *fetch_channel()*

Кожне повідомлення, що пройшло перевірку на наявність тексту та очищення, записується до бази даних.

- **Паралельне виконання та збереження результатів**

Після цього головна функція *main()* запускає збір даних з усіх п'яти каналів одночасно за допомогою *asyncio.gather()*, що об'єднує всі асинхронні задачі та виконує їх конкурентно. Після завершення всіх задач викликається *conn.commit()* для фіксації транзакцій у базі даних та *conn.close()* для коректного закриття з'єднання. У завершальній частині програма виводить у термінал статистику: кількість збережених повідомлень по кожному каналу окремо та загальний підсумок (рис. 2.2.9).

```

Збір даних завершено!
Zelenskiy / Official: 256 повідомлень
Лачен пише: 1696 повідомлень
Труха ⚡ Україна: 2360 повідомлень
Телеграмна служба новин - Україна: 2497 повідомлень
Реальна Війна | Україна | Новини: 3143 повідомлень
Загалом: 9952 повідомлень

```

Рисунок 2.2.9. Результати виконання програми збору даних

- **Структура бази даних**

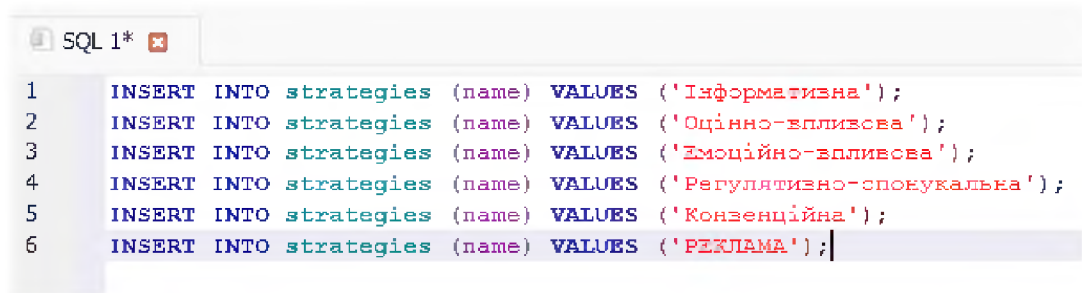
Для перегляду та подальшої роботи з базою даних використовувався SQL-переглядач DB Browser for SQLite (DB Browser for SQLite, n. d.).

База даних реалізує реляційну модель із чотирьох таблиць. Таблиця **messages** зберігає основний корпус текстів і містить чотири поля: *id* (числовий ідентифікатор повідомлення в Telegram), «Джерело» (повна текстова назва каналу), «Дата публікації» (дата і час у форматі YYYY-MM-DD HH:MM:SS) та «Текст повідомлення» (відфільтрований текст). Первинний ключ є складеним і включає поля *id* та «Джерело», що гарантує унікальність кожного запису навіть у разі збігу числових ідентифікаторів між різними каналами. Директива INSERT OR IGNORE запобігає дублюванню даних при повторному запуску програми (рис. 2.2.10).

Таблиця: messages					Filter in any column
	id	Джерело	Дата публікації	Текст повідомлення	
	Фільтр	Фільтр	Фільтр	Фільтр	
61	82750	Реальна Війна Україна Новини	2024-01-30 22:39:38	Є і деякі законні причини неяви до ...	
62	82749	Реальна Війна Україна Новини	2024-01-30 22:22:06	Нацполіція, ДВР, чиновники та подібн...	
63	82746	Реальна Війна Україна Новини	2024-01-30 22:12:20	Як відбувається реабілітація ...	
64	82744	Реальна Війна Україна Новини	2024-01-30 22:06:50	Поліцію теж задіють для мобілізації:...	
65	82743	Реальна Війна Україна Новини	2024-01-30 22:02:38	Чоловікам за кордоном: Будь-які ...	
66	82742	Реальна Війна Україна Новини	2024-01-30 21:51:34	Жінки: щодо придатних, які мають ...	
67	82741	Реальна Війна Україна Новини	2024-01-30 21:44:16	Обмежено-придатні проходять ...	
68	82740	Реальна Війна Україна Новини	2024-01-30 21:42:04	Ті, хто засуджені на умовний термін ...	
69	82739	Реальна Війна Україна Новини	2024-01-30 21:39:32	Строкова військова служба замінена ...	
70	82738	Реальна Війна Україна Новини	2024-01-30 21:27:46	Не підлягаєть призову на військову ...	
71	82737	Реальна Війна Україна Новини	2024-01-30 21:22:56	Виключена норма захисту від призову ...	
72	82736	Реальна Війна Україна Новини	2024-01-30 21:18:27	Обов'язково мати при собі військово-...	
73	82735	Реальна Війна Україна Новини	2024-01-30 21:16:00	Повістки тепер приходять також й...	
74	82734	Реальна Війна Україна Новини	2024-01-30 21:09:20	Призовний вік з 25 років, обмеження ...	
75	82732	Реальна Війна Україна Новини	2024-01-30 21:04:23	В мордорі горять ТЗС-23, москвічі ...	
76	38875	Лачен пише	2024-01-31 23:54:45	раhistи також обстріляли авіабомбами...	
77	38874	Лачен пише	2024-01-31 23:25:05	Попередньо є влучання по цивільній ...	

Рисунок 2.2.10. Вигляд таблиці *messages* у БД

Таблиця **strategies** є довідником стратегій: кожна стратегія зберігається в ній рівно один раз. Створення таблиці передбачене програмним кодом, а записи в неї були додані вручну в процесі анотування корпусу (рис. 2.2.11). Ця таблиця містить поля *id* (автоінкрементний первинний ключ) та *name* (унікальна текстова назва стратегії) (рис. 2.2.12).

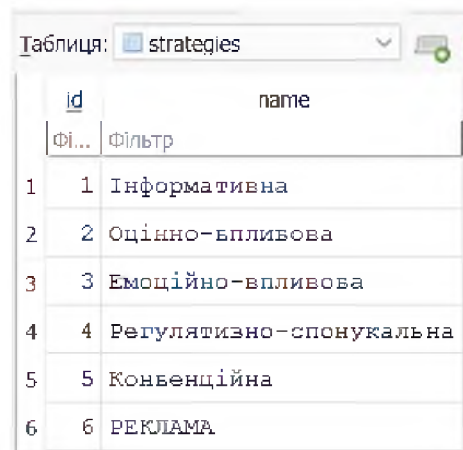


```

1  INSERT INTO strategies (name) VALUES ('Інформативна');
2  INSERT INTO strategies (name) VALUES ('Оцінно-впливова');
3  INSERT INTO strategies (name) VALUES ('Емоційно-впливова');
4  INSERT INTO strategies (name) VALUES ('Регулятивно-спонукальна');
5  INSERT INTO strategies (name) VALUES ('Конвенційна');
6  INSERT INTO strategies (name) VALUES ('РЕКЛАМА');

```

Рисунок 2.2.11. Наповнення таблиці *strategies* через SQL-запити



Таблиця: strategies	
id	name
1	Інформативна
2	Оцінно-впливова
3	Емоційно-впливова
4	Регулятивно-спонукальна
5	Конвенційна
6	РЕКЛАМА

Рисунок 2.2.12. Вигляд таблиці *strategies* у БД

Таблиця **sources** є довідником каналів і містить два поля: *id* (числовий ідентифікатор для кожного джерела від 1 до 5) та *name* (повна назва каналу). Вона була створена та наповнена вручну через SQL-запити (рис. 2.13, 2.14). Вона слугує сполучною ланкою між текстовими назвами каналів у таблиці *messages* та числовими ідентифікаторами у проміжній таблиці.

```

1 CREATE TABLE sources (
2     id INTEGER PRIMARY KEY AUTOINCREMENT,
3     name TEXT UNIQUE NOT NULL
4 );
5
6
7 INSERT INTO sources (name)
8 SELECT DISTINCT "Джерело" FROM messages;
9
10
11 DROP TABLE message_strategies;
12
13 CREATE TABLE message_strategies (
14     message_id INTEGER NOT NULL,
15     source_id INTEGER NOT NULL REFERENCES sources(id),
16     strategy_id INTEGER NOT NULL REFERENCES strategies(id),
17     PRIMARY KEY (message_id, source_id, strategy_id)
18 );

```

Рисунок 2.2.13. Створення та наповнення таблиці *sources*

Таблиця: sources

	id	name
1	1	Zelenskiy / Official
2	2	Лачен пише
3	3	Телеграмна служба новин - Україна
4	4	Реальна Війна Україна Новини
5	5	Труха ⚡ Україна

Рисунок 2.2.14. Вигляд готової таблиці *sources* у БД

Таблиця **message_strategies** є проміжною і реалізує зв'язок «багато-до-багатьох» між повідомленнями та стратегіями. Вона містить три поля: *message_id*, *source_id* та *strategy_id*, що разом утворюють складений первинний ключ. Завдяки такій структурі одне повідомлення може бути пов'язане з кількома стратегіями одночасно, що відповідає реальній природі досліджуваного матеріалу. Зовнішні ключі забезпечують цілісність даних, а їхня підтримка активована директивою `PRAGMA foreign_keys = ON`. Початково ця таблиця порожня для подальшої ручної та автоматизованої розмітки.

Таким чином, загальна структура створеної бази даних має вигляд зображений на рисунку 2.2.15, а повна БД розміщена у GitHub репозиторії (див. Додаток 1).

Ім'я	Тип
▼ Таблиці (5)	
▼ message_strategies	
message_id	INTEGER
source_id	INTEGER
strategy_id	INTEGER
▼ messages	
id	INTEGER
Джерело	TEXT
Дата публікації	TEXT
Текст повідомлення	TEXT
▼ sources	
id	INTEGER
name	TEXT
> sqlite_sequence	
▼ strategies	
id	INTEGER
name	TEXT

Рисунок 2.2.15. Загальна структура бази даних

Отже, розроблена програма забезпечує повний цикл автоматизованого збору та первинної обробки текстових даних із п'яти україномовних Telegram-каналів. Завдяки застосуванню асинхронної бібліотеки Telethon, вдалося організувати паралельне збирання дописів у межах заданого часового інтервалу, що помітно збільшує продуктивність даного процесу. Багаторівнева процедура мовного очищення та лінгвістичного сортування гарантує високу якість та єдиномовність кінцевого корпусу, фільтруючи рекламні посилання, службові позначки, а також тексти, що є російськомовними або повністю англomовними. Сформована реляційна бази даних із чотирьох таблиць, реалізована засобами SQLite, не лише зберігає сформований корпус, а й закладає архітектурну основу для подальшого етапу анотування: довідники стратегій і джерел у поєднанні з проміжною таблицею зв'язків забезпечують гнучке і структуроване маркування вербальних тактик маніпуляції без надлишкового дублювання даних.

Висновки до другого розділу

Другий розділ бакалаврської роботи присвячено формуванню емпіричної бази дослідження та її програмно-технічному забезпеченню. У підрозділі 2.1 визначено методологічну основу роботи, яку складають два взаємодоповнювальні методи: контент-аналіз і дискурс-аналіз. Перший забезпечує кількісне вимірювання повторюваних мовних тенденцій, другий –

розкриває прихований зміст повідомлень у соціально-культурному контексті воєнного часу. Обґрунтовано вибір платформи Telegram як основного середовища дослідження, що зумовлено її домінуючою роллю у новинному просторі України в умовах повномасштабного вторгнення. Сформовану вибірку складають п'ять каналів різних типів – анонімні агрегатори новин, офіційний державний канал та авторський канал із верифікованим автором. Це забезпечує типологічну різноманітність матеріалу та уможливлює порівняльний аналіз вербальних стратегій різних суб'єктів медіапростору. Хронологічні межі дослідження – грудень 2023 – січень 2024 року – визначено як оптимальні, оскільки цей період поєднує мовну однорідність корпусу з характерною для воєнного часу насиченістю стратегіями впливу.

У підрозділі 2.2 описано програмну реалізацію системи автоматизованого збору та первинної обробки даних. Розроблений скрипт на мові Python із використанням бібліотеки Telethon забезпечує паралельне вивантаження повідомлень із п'яти каналів у межах заданого часового діапазону. Багатоетапна функція лінгвістичного очищення та мовного відбору гарантує якість і мовну однорідність сформованого корпусу, усуваючи непотрібні для корпусу URL-посилання та рекламні вставки, технічні артефакти, російськомовні та повністю англomовні тексти. Результатом роботи програми стала структурована реляційна база даних SQLite обсягом 9952 повідомлення, архітектура якої складається з чотирьох таблиць: основного корпусу текстів, довідників каналів і стратегій та проміжної таблиці зв'язків. Така структура не лише зберігає зібраний матеріал, а й закладає готову основу для подальшого етапу дослідження – ручної розмітки та автоматизованої класифікації вербальних стратегій, яким присвячено третій розділ роботи.

РОЗДІЛ 3. ЕКСПЕРТНА РОЗМІТКА ТА АВТОМАТИЗОВАНА КЛАСИФІКАЦІЯ МЕТОДАМИ МАШИННОГО НАВЧАННЯ

У цьому розділі на основі теоретичних засад та емпіричної бази, сформованих у попередніх розділах, реалізовано практичну частину дослідження: здійснено експертну розмітку корпусу Telegram-повідомлень, побудовано систему автоматизованої класифікації вербальних стратегій засобами машинного навчання та проаналізовано отримані результати. Розділ складається з чотирьох підрозділів: перший присвячено принципам лінгвістичного анотування та формуванню тренувального датасету, другий – архітектурі та навчанню моделі класифікації, третій – автоматизованій розмітці повного корпусу, четвертий – інтерпретації результатів та порівняльному аналізу стратегій у розрізі каналів.

3.1. Принципи лінгвістичного анотування та формування тренувального датасету

За основу класифікаційної схеми взято типологію комунікативних стратегій Т. Толмачової (Маслова, н. д., с. 83), адаптовану до специфіки новинного Telegram-дискурсу воєнного часу. Відповідно до цієї типології, у процесі ручної розмітки корпусу використовувалися п'ять стратегій: інформативна, оцінно-впливова, емоційно-впливова, регулятивно-спонукальна та конвенційна. Додатково було введено мітку РЕКЛАМА, якою позначалися повідомлення комерційного характеру, що не несуть новинної чи дискурсивної цінності для дослідження. До цієї категорії відносилися як очевидна реклама товарів і послуг, так і заклики підписатися на інші Telegram-канали, рекламні оголошення курсів, акцій тощо (рис. 3.1.1).

НОВОРІЧНИЙ БУМ ЦІН... ...на рекламу в мережі каналів «Труха»!
Встигни отримати КРАЩУ ЦІНУ РОКУ на просування свого бізнесу чи
бренду. Для цього: Напиши менеджеру рекламної агенції
PromotionManagerTo Обговори деталі Відправ рекламний пост Отримай
величезну кількість клієнтів! ЗАЛИШИЛАСЬ ОБМЕЖЕНА КІЛЬКІСТЬ МІСЦЬ –
ВСТИГНИ!

В Новий рік – з новими смартфонами ZTE! Купуйте новинки серії ZTE
Blade за святковими цінами від 3999 грн! ZTE Blade A54 4/128Gb за
спеціальною ціною 3999 грн 6.6" HD дисплей / Батарея 5000 мАгод /
RAM 44 GB ZTE Blade A73 4/128Gb за 4699 грн 6.6" 90 Гц дисплей /
основна камера на 50 Мп / RAM 44 GB ZTE Blade V50 Vita 6/128Gb за
спеціальною ціною 4899 грн 6.75" 90 Гц дисплей / основна камера на
50 Мп / RAM 66 GB ZTE Blade V50 Design 8/128Gb за спеціальною ціною
від 5699 грн 6.6" FHD дисплей / основна камера на 50 Мп / RAM 88 GB
Вільше крутих пропозицій на каналі АЛЛО реклама

Рисунок 3.1.1. Приклад рекламних повідомлень

Введення цієї позначки зумовлене специфічною природою рекламних повідомлень: такі тексти нерідко імітують стилістику звичайних новинних дописів, використовуючи схожі мовні конструкції та риторичні прийоми, що ускладнює їх однозначне віднесення до будь-якої з п'яти основних стратегій. Саме тому вони були виокремлені в окрему категорію як самостійний тип комунікативної поведінки, що становить інтерес для аналізу медіадискурсу. Важливо зазначити, що мітка РЕКЛАМА є взаємовиключною з іншими категоріями: якщо повідомленню присвоювалась ця мітка, подальше визначення додаткових стратегій не здійснювалося.

Метою етапу ручної розмітки було сформувати тренувальний датасет обсягом близько 1000 анотованих повідомлень із представленими прикладами для кожної з шести категорій.

- **Операційні критерії розмітки**

Визначення стратегії здійснювалося на основі домінантної комунікативної функції повідомлення. Одному повідомленню могло бути присвоєно кілька міток, якщо в ньому поєднувалися функції різних стратегій.

Інформативна стратегія ідентифікувалася за наявністю нейтральної фактологічної подачі матеріалу: повідомлення про події, їх учасників, час і місце. Характерні маркери – безособові конструкції, дієслова стану та руху у минулому часі, числові дані, географічні назви («повідомляють», «за даними», «станом на», «загинули», «відбулось»).

Оцінно-впливова стратегія визначалася за наявністю явних або прихованих оцінних суджень, аксіологічно забарвленої лексики, риторичних запитань та апеляцій до цінностей аудиторії («злочин», «ганебно», «герої», «зрада», «справжній», «неприйнятно»).

Емоційно-впливова стратегія фіксувалася у повідомленнях, спрямованих на збудження або зміну психоемоційного стану читача: використання вигуків, емодзі, патетичної лексики, риторичних звертань, синтаксичних конструкцій із посиленням («Тримайтесь!», «Неймовірно», «Це треба бачити», «ЖАХ»).

Регулятивно-спонукальна стратегія ідентифікувалася за наявністю спонукальних конструкцій, що апелюють до дій читача: прямі заклики, поради, прохання, вимоги («задонатйте», «зверніться», «не виходьте», «повідомте», «підтримайте»).

Конвенційна стратегія фіксувалася у повідомленнях, що виконують функцію підтримання контакту з аудиторією або організації комунікативного простору каналу: привітання, подяки, звернення до підписників, анонси («Вітаю всіх!», «Дякуємо за підтримку», «Слідкуйте за оновленнями»).

- **Інструмент ручної розмітки**

Для спрощення процедури анотування розроблено допоміжний Python-скрипт із інтерактивним інтерфейсом у терміналі (рис. 3.1.2), який також завантажено до загального GitHub репозиторію (див. Додаток 1).

```
Стратегії:
1 – Інформативна
2 – Оцінно-впливова
3 – Емоційно-впливова
4 – Регулятивно-спонукальна
5 – Конвенційна
6 – РЕКЛАМА

Команди:
1,2,3 – номери стратегій через кому
Enter – пропустити (розмітити пізніше)
d – видалити повідомлення з БД
q – вийти, зберігши прогрес

Залишилось розмітити: 9535 повідомлень
```

Рисунок 3.1.2. Фрагмент інтерфейсу скрипту для ручного анотування

Програма послідовно виводить повідомлення з бази даних у випадковому порядку, відображаючи лише ті записи, яким ще не присвоєно жодної стратегії. Перед кожним повідомленням виводиться актуальна статистика розмітки у вигляді текстового графіка (рис. 3.1.3).

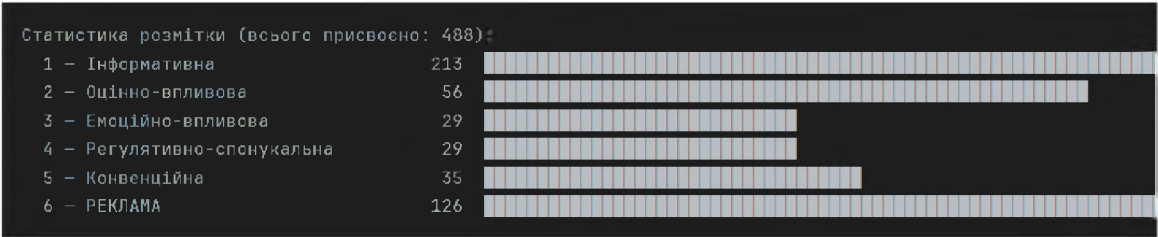


Рисунок 3.1.3. Графічна статистика ручної розмітки

Передбачено також можливість пропустити повідомлення для пізнішого розгляду, видалити його з бази даних або зупинити роботу зі збереженням прогресу (рис. 3.1.4).

```
=====
Джерело: Реальна Війна | Україна | Новини | Дата: 2024-01-04 11:05:01
Текст:
4 січня на аеродромі "Шагол" у російському Челябінську згорів винищувач-
бомбардувальник Су-34 Ціна винищувача – 36 млн доларів.
=====
Введи номери стратегій / d щоб видалити / q щоб вийти / Enter щоб пропустити: 1
```

Рисунок 3.1.4. Процес ручного анотування повідомлень

На цьому етапі я виступала в ролі експерта та зазначала номери відповідних стратегій через кому, після чого вони автоматично записувалися до таблиці *message_strategies* (рис. 3.1.5).

Таблиця: message_strategies

	<u>message_id</u>	<u>source_id</u>	<u>strategy_id</u>
	Фільтр	Фільтр	Фільтр
1	84661	5	1
2	38530	2	1
3	74143	3	1
4	81501	4	1
5	8916	1	1
6	8916	1	5
7	8916	1	3
8	86148	5	3

Рисунок 3.1.5. Автоматичне заповнення таблиці *message_strategies* у БД

- **Важливість порівняльного аналізу різних каналів**

Показовим аспектом дослідження є те, що одна й та сама подія могла висвітлюватися принципово по-різному залежно від каналу-джерела. Відмінності стосувалися не лише тону чи лексичного наповнення, а й домінантної стратегії: там, де один канал обмежувався нейтральною інформативною подачею, інший вдавався до емоційно-впливових або оцінних конструкцій. Це підтверджує доцільність формування вибірки з п'яти типологічно різних каналів і робить порівняльний аналіз розподілу стратегій між ними окремо значущим (рис. 3.1.6).

Реальна Війна Україна Новини	2... 1...	В Сімферополі дві диверсантки отруїли військових – кацапський тт-канал Повідомляють, що дівчата пригостили їх їжею та горілкою, куди додали з миш'як і стрихнін. У результаті такої трапези 24 особи загинуло та ще 11 у лікарні.
Телеграмна служба новин – Україна	2... 1...	Дві диверсантки отруїли російських військових у Сімферополі, – росЗМІ Повідомляють, що дівчата пригостили їх їжею та горілкою з миш'яком та стрихніном. В результаті 24 особи загинуло та ще 11 у лікарні.

Рисунок 3.1.6. Приклад відмінної подачі однієї новини різними каналами

- **Труднощі ручної розмітки**

У процесі анотування виявлено кілька категорій проблемних випадків, які потребували ручного втручання. По-перше, у незначній кількості повідомлень наприкінці тексту залишалися нефільтровані рекламні вставки із закликом до підписки або до співпраці з каналом – нестандартні формулювання, не передбачені регулярними виразами попереднього етапу обробки (рис. 3.1.7).

```
Міноборони розірвало контракти з
компаніями Гринкевича. Окрім того,
міністерство подало в ДВР заяву про
визнання потерпілою стороною у справі
проти компаній бізнесмена.
ТРУХАУкраїна Прислати новину
```

Рисунок 3.1.7. Приклад повідомлення, що не пройшло фільтрацію

По-друге, до корпусу потрапила невелика кількість повністю російськомовних текстів, у яких були відсутні характерні маркерні літери (ы, э, ъ, ё), що ускладнювало їхнє автоматичне відсіювання. Усі виявлені проблемні записи було видалено вручну безпосередньо в процесі анотування (рис. 3.1.8).

85409	Труха 🇺🇦 Україна	2024-01-08 07:55:56	Все 4 борта МиГ-31К совершили посадку в Саваслейке, – Ванек.
85379	Труха 🇺🇦 Україна	2024-01-08 07:14:44	Первая волна ракет в сторону Шепетовского р-на Хмельницкой области, вероятно курсом на Житомирскую область, – Ванек.
84771	Труха 🇺🇦 Україна	2023-12-29 09:21:11	Еще пуски через Сумскую область вектором на Полтавську/Черкаську області, – Ванек.
84709	Труха 🇺🇦 Україна	2023-12-29 06:59:06	Валлистика в сторону Днепропетровской области, – Ванек.

Рисунок 3.1.8. Приклад текстів, що видалялися вручну

Після завершення ручної фільтрації проблемних записів та анотування потрібної кількості стратегій, загальний обсяг корпусу становив 9939 одиниць. За результатами розмітки було присвоєно 1540 стратегій загалом, з них: інформативна – 497, оцінно-впливова – 245, емоційно-впливова – 223, регулятивно-спонукальна – 201, конвенційна – 200, РЕКЛАМА – 174. Нерівномірний розподіл між класами зумовлений об'єктивною природою досліджуваного матеріалу: інформативна стратегія є домінантною в новинному дискурсі і здебільшого вживається у поєднанні з іншими стратегіями, тоді як конвенційна, регулятивно-спонукальна чи РЕКЛАМА функціонально обмежені й трапляються значно рідше. Отриманий розподіл відображає реальну частотність стратегій в українському новинному Telegram-дискурсі воєнного часу, що саме по собі є результатом дослідження.

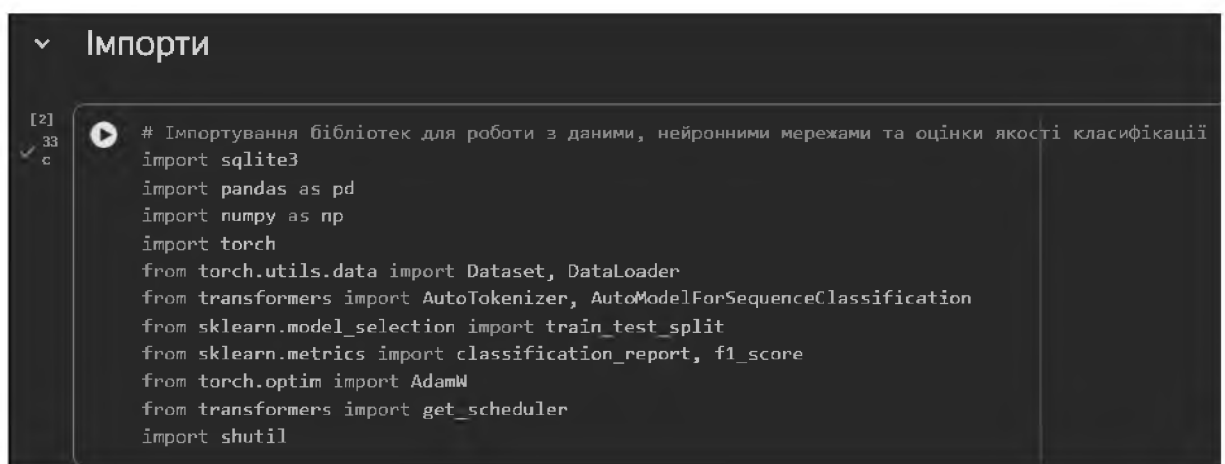
Таким чином, процес ручного анотування корпусу ґрунтувався на типології Т. Толмачової та передбачав присвоєння кожному повідомленню однієї або кількох міток із шести категорій з орієнтиром на збалансований розподіл прикладів. Розроблений допоміжний скрипт суттєво спростив процедуру розмітки, забезпечивши зручний інтерактивний інтерфейс та автоматичне збереження результатів до бази даних. Виявлені в процесі анотування проблемні записи було усунено вручну, що підвищило якість і однорідність фінального датасету обсягом 9939 повідомлень. Загалом розмічена вибірка охопила 908 повідомлень із 1540 присвоєними мітками стратегій, що слугує тренувальною основою для подальшої автоматизованої класифікації стратегій засобами машинного навчання.

3.2. Архітектура, налаштування та навчання моделі класифікації вербальних стратегій

Для автоматизованої класифікації вербальних стратегій у корпусі Telegram-повідомлень розроблено систему машинного навчання на основі попередньо навченої багатомовної мовної моделі. Повний програмний код реалізовано в середовищі Google Colab (Google, n. d.) та розміщено у Додатку 2.

- **Використані бібліотеки та їхнє призначення**

Реалізацію системи виконано мовою програмування Python із використанням декількох спеціалізованих бібліотек (рис. 3.2.1).



```
▼ Імпорти

[2]: # Імпортування бібліотек для роботи з даними, нейронними мережами та оцінки якості класифікації
import sqlite3
import pandas as pd
import numpy as np
import torch
from torch.utils.data import Dataset, DataLoader
from transformers import AutoTokenizer, AutoModelForSequenceClassification
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report, f1_score
from torch.optim import AdamW
from transformers import get_scheduler
import shutil
```

Рисунок 3.2.1. Імпортування бібліотек для реалізації системи класифікації

Центральною є бібліотека *transformers* від компанії Hugging Face (Hugging Face, n. d.-a), яка надає доступ до попередньо навчених трансформерних моделей, зокрема BERT, і забезпечує інструменти для їхнього донавчання на власних даних. Бібліотека *torch* (PyTorch) реалізує обчислювальний граф нейронної мережі, механізм зворотного поширення помилки та підтримку GPU-обчислень. Бібліотека *scikit-learn* використовується для розподілу даних на тренувальну і валідаційну вибірки та обчислення метрик якості класифікації. Бібліотека *pandas* забезпечує зручну роботу з табличними даними при завантаженні корпусу з бази даних, а *numpy* – числові операції над масивами векторів і міток.

- **Вибір архітектури моделі**

За основу класифікатора обрано модель **bert-base-multilingual-cased** (Hugging Face, n. d.-b) – багатомовну версію архітектури BERT (Bidirectional Encoder Representations from Transformers), яка підтримує 104 мови, включно з українською, і не потребує окремого україномовного корпусу для попереднього навчання. На відміну від традиційних підходів на основі TF-IDF або word2vec, BERT враховує двобічний контекст кожного слова в реченні, що дозволяє розрізняти семантично близькі конструкції – наприклад, нейтральну фактологічну подачу від емоційно забарвленої.

Задача класифікації сформульована як **multi-label** – багатомітковий варіант, при якому одному повідомленню може бути присвоєно кілька стратегій одночасно. Це відповідає реальній природі досліджуваного матеріалу: повідомлення нерідко поєднують, наприклад, інформативну та оцінно-впливову стратегії в межах одного тексту. Для реалізації multi-label класифікації до базової архітектури BERT додано лінійний класифікаційний шар із шістьма виходами – по одному на кожну стратегію. На виході застосовується функція *sigmoid* (а не softmax, як у звичайній одноміткової класифікації), яка дозволяє кожному класу незалежно отримувати значення від 0 до 1, що інтерпретується як ймовірність присутності відповідної стратегії.

- **Завантаження та підготовка даних**

На першому етапі розмічені повідомлення завантажуються з реляційної бази даних SQLite за допомогою SQL-запиту, який об'єднує таблиці повідомлень і стратегій та формує для кожного запису рядок із переліком присвоєних міток. Загалом завантажено 908 розмічених повідомлень (рис. 3.2.2).

```

... Всього розмічених повідомлень: 908
strategies
6      174
1      162
1,2    72
4      55
1,5    54
1,3    47
1,4    42
2,3    31
3      29
1,2,3  29
Name: count, dtype: int64

```

Рисунок 3.2.2. Результати виконання SQL-запиту

На другому етапі кожному повідомленню присвоюється бінарний вектор міток довжиною 6, де кожна позиція відповідає одній із стратегій: значення 1 означає присутність стратегії, 0 – її відсутність. Наприклад, повідомлення з мітками «інформативна» та «оцінно-впливова» отримує вектор [1, 1, 0, 0, 0, 0].

Корпус розподіляється на тренувальну (80%) і валідаційну (20%) вибірки методом випадкового розбиття зі фіксованим станом генератора випадкових чисел для відтворюваності результатів. Таким чином, тренувальна вибірка налічує 726 повідомлення, валідаційна – 182.

- **Токенізація**

Перед подачею в модель тексти проходять токенизацію за допомогою токенизатора `bert-base-multilingual-cased`. Кожне повідомлення перетворюється на послідовність числових ідентифікаторів токенів з фіксованою максимальною довжиною 256 одиниць: коротші тексти доповнюються спеціальним токеном заповнення [PAD], довші – обрізаються. Втім, з огляду на характерну для Telegram-дискурсу стислість повідомлень, це обмеження не є критичним для більшості одиниць корпусу. Одночасно формується маска уваги (*attention mask*), яка вказує моделі які позиції є значущими, а які є заповненням і мають ігноруватись. Реалізацію цього етапу забезпечує клас `TelegramDataset`, успадкований від базового класу `Dataset` бібліотеки *PyTorch* (рис. 3.2.3).

```

class TelegramDataset(Dataset):
    """
    Клас датасету для подачі текстових повідомлень у модель BERT.
    Кожне повідомлення токенизується і доповнюється до фіксованої довжини 256 токенів.
    """
    def __init__(self, texts, labels, tokenizer, max_length=256):
        self.texts = texts
        self.labels = labels
        self.tokenizer = tokenizer
        self.max_length = max_length

    def __len__(self):
        return len(self.texts)

    def __getitem__(self, idx):
        encoding = self.tokenizer(
            self.texts[idx],
            max_length=self.max_length,
            padding='max_length',
            truncation=True,
            return_tensors='pt'
        )
        return {
            'input_ids': encoding['input_ids'].squeeze(),
            'attention_mask': encoding['attention_mask'].squeeze(),
            'labels': torch.tensor(self.labels[idx], dtype=torch.float)
        }

```

Рисунок 3.2.3. Реалізація класу TelegramDataset для токенизації та підготовки даних

- **Налаштування навчання та функція втрат**

Навчання моделі здійснюється протягом 6 епох з використанням оптимізатора AdamW зі швидкістю навчання $2e-5$, яка є стандартною для задач fine-tuning BERT. Застосовується лінійний планувальник швидкості навчання (linear scheduler) з 50 кроками розігріву, що запобігає різким стрибкам градієнтів на початку навчання.

Оскільки розподіл стратегій у корпусі є нерівномірним – інформативна стратегія представлена значно більше за інші, – для компенсації дисбалансу класів використано зважену функцію втрат *BCEWithLogitsLoss* з автоматично розрахованими коефіцієнтами *pos_weight*. Ваговий коефіцієнт для кожного класу обчислюється як відношення кількості негативних прикладів до кількості позитивних: класи з меншою кількістю представників отримують вищу вагу в функції втрат, що змушує модель приділяти їм більше уваги під час навчання (рис. 3.2.4).

```

EPOCHS = 6

labels_matrix_train = np.array(y_train)
pos_counts = labels_matrix_train.sum(axis=0)
neg_counts = len(y_train) - pos_counts
pos_weight = torch.tensor(neg_counts / (pos_counts + 1e-5), dtype=torch.float).to(device)

loss_fn = torch.nn.BCEWithLogitsLoss(pos_weight=pos_weight)
optimizer = AdamW(model.parameters(), lr=2e-5)

num_training_steps = EPOCHS * len(train_loader)
lr_scheduler = get_scheduler(
    'linear',
    optimizer=optimizer,
    num_warmup_steps=50,
    num_training_steps=num_training_steps
)

```

Рисунок 3.2.4. Налаштування функції втрат та оптимізатора

Безпосередній процес навчання реалізовано у вигляді циклу по епохах, на кожній ітерації якого модель отримує батч токенизованих текстів, обчислює функцію втрат і оновлює ваги через зворотне поширення помилки (рис. 3.2.5).

```

for epoch in range(EPOCHS):
    # Фаза навчання
    model.train()
    total_loss = 0
    for batch in train_loader:
        input_ids = batch['input_ids'].to(device)
        attention_mask = batch['attention_mask'].to(device)
        labels_batch = batch['labels'].to(device)

        outputs = model(input_ids=input_ids, attention_mask=attention_mask)
        loss = loss_fn(outputs.logits, labels_batch)

        optimizer.zero_grad()
        loss.backward()
        optimizer.step()
        lr_scheduler.step()

        total_loss += loss.item()

    avg_loss = total_loss / len(train_loader)

```

Рисунок 3.2.5. Цикл навчання моделі по епохах

Навчання відбувається на GPU, наявність якого визначається автоматично засобами PyTorch. Застосування GPU середовища Google Colab скорочує час однієї епохи до кількох хвилин.

- **Підбір оптимальних порогів класифікації**

Оскільки модель на виході видає не бінарні мітки, а неперервні значення від 0 до 1, необхідно встановити поріг, вище якого стратегія вважається присутньою. Замість фіксованого порогу 0.5, який є стандартним за замовчуванням, для кожного класу підбирається індивідуальний оптимальний поріг у діапазоні від 0.2 до 0.7 з кроком 0.05. Критерієм вибору слугує максимальний F1-score на валідаційній вибірці. Такий підхід дозволяє краще балансувати між точністю і повнотою для класів із різною частотністю в корпусі (рис. 3.2.6).

```
print("Оптимальні пороги класифікації по стратегіях:")
best_thresholds = []
for i in range(NUM_LABELS):
    best_f1 = 0
    best_t = 0.5
    for t in np.arange(0.3, 0.75, 0.05):
        preds = (all_probs[:, i] > t).astype(int)
        f1 = f1_score(all_labels_raw[:, i], preds, zero_division=0)
        if f1 > best_f1:
            best_f1 = f1
            best_t = t
    best_thresholds.append(best_t)
print(f" {strategy_names[i]}: porir={best_t:.2f}, F1={best_f1:.3f}")
```

Рисунок 3.2.6. Реалізація алгоритму підбору оптимальних порогів класифікації

За результатами підбору встановлено такі оптимальні пороги: інформативна – 0.35, оцінно-впливова – 0.40, емоційно-впливова – 0.55, регулятивно-спонукальна – 0.50, конвенційна – 0.65, РЕКЛАМА – 0.70. Нижчі пороги для інформативної та оцінно-впливової стратегій свідчать про те, що модель поводить себе обережно з цими класами і потребує меншої впевненості для їхнього присвоєння, тоді як вищий поріг для, наприклад, конвенційної відображає більшу складність її розпізнавання.

- **Оцінка якості та інтерпретація результатів**

Якість класифікації оцінюється за трьома метриками: precision (точність) – частка правильно присвоєних міток серед усіх присвоєних; recall (повнота) – частка правильно знайдених стратегій серед усіх реальних; F1-score – їхнє гармонічне середнє, яке враховує обидва аспекти одночасно. Для задачі

виявлення вербальних стратегій пріоритетнішою є повнота: краще знайти зайву стратегію, ніж пропустити реальну (рис. 3.2.7).

Зведений звіт якості класифікації:				
	precision	recall	f1-score	support
Інформативна	0.88	0.94	0.91	113
Оцінно-впливова	0.48	0.81	0.60	47
Емоційно-впливова	0.57	0.66	0.61	44
Регулятивно-спонукальна	0.54	0.69	0.61	39
Конвенційна	0.89	0.85	0.87	40
РЕКЛАМА	0.82	1.00	0.90	28
micro avg	0.70	0.84	0.77	311
macro avg	0.70	0.82	0.75	311
weighted avg	0.73	0.84	0.78	311
samples avg	0.77	0.87	0.79	311

Рисунок 3.2.7. Зведений звіт якості класифікації

Зведений звіт якості класифікації з оптимальними порогоми демонструє такі результати по кожній стратегії:

- найвищий F1-score отримала **інформативна** стратегія – 0.91, що зумовлено найбільшою кількістю тренувальних прикладів і формально впізнаваними конструкціями нейтральної фактологічної подачі.
- Категорія **РЕКЛАМА** досягла F1-score 0.90 з показником повноти 1.00 – модель не пропустила жодного рекламного повідомлення, що пояснюється лексичною специфічністю цього класу: рекламні тексти містять характерні маркери – комерційну лексику, заклики до купівлі, цінові пропозиції, – які чітко відрізняють їх від інших типів.
- **Конвенційна** стратегія показала F1-score 0.87 – високий результат, що пояснюється функціональною чіткістю цього класу: привітання, подяки та звернення до підписників реалізуються через стійкі етикетні формули, які модель впевнено розпізнає.
- **Оцінно-впливова, емоційно-впливова та регулятивно- спонукальна** стратегії отримали F1-score у діапазоні 0.60–0.61. Нижчі показники цих класів мають очевидне лінгвістичне пояснення: зазначені стратегії семантично близькі між собою і нерідко реалізуються через схожі мовні засоби – оцінну лексику, спонукальні конструкції, емоційно забарвлені

звертання. Їхнє розмежування становить складність навіть для людини-анотатора, тому певний рівень помилок моделі є закономірним.

Загальний показник F1 macro склав 0.75, що є гідним результатом для задачі multi-label класифікації на корпусі обсягом близько 908 повідомлень із шістьма класами. Показово, що recall по всіх класах є вищим за precision – модель схильна радше присвоїти зайву мітку, ніж пропустити стратегію, що відповідає дослідницьким пріоритетам аналізу медіадискурсу.

Таким чином, розроблена система класифікації на основі fine-tuning моделі bert-base-multilingual-cased забезпечує автоматичне розпізнавання шести типів вербальних стратегій у новинних Telegram-повідомленнях із загальним F1-score 0.75. Диференційована якість по класах відображає реальну лінгвістичну складність задачі: лексично специфічні категорії (РЕКЛАМА, інформативна, конвенційна) розпізнаються значно краще за семантично близькі (оцінно-впливова, емоційно-впливова, регулятивно-спонукальна). Навчена і збережена модель слугує основою для подальшої автоматизованої розмітки всього корпусу, що є предметом наступного підрозділу.

3.3. Автоматизована розмітка корпусу

На завершальному етапі навчена модель застосовується для автоматичної класифікації вербальних стратегій у повному корпусі нерозмічених повідомлень. Програмний код цього етапу також реалізовано в середовищі Google Colab та розміщено у Додатку 3.

На першому кроці навчена модель і токенизатор завантажуються з Google Drive, де вони були збережені після навчання (рис. 3.3.1).

```
from google.colab import drive
drive.mount('/content/drive')

from transformers import AutoTokenizer, AutoModelForSequenceClassification
import torch

MODEL_PATH = '/content/drive/MyDrive/strategy_classifier'
device = torch.device('cuda' if torch.cuda.is_available() else 'cpu')

tokenizer = AutoTokenizer.from_pretrained(MODEL_PATH)
model = AutoModelForSequenceClassification.from_pretrained(MODEL_PATH)
model = model.to(device)
model.eval()
print("Модель завантажена")

Mounted at /content/drive
Loading weights: 100% 201/201 [00:04<00:00,
Модель завантажена
```

Рисунок 3.3.1. Блок завантаження моделі

Одночасно з Drive завантажуються база даних корпусу, з якої за допомогою SQL-запиту вибираються всі повідомлення, що не мають жодної присвоєної стратегії – тобто ті, що не увійшли до тренувальної вибірки та не були розмічені вручну (рис. 3.3.2).

```
import sqlite3
import pandas as pd

# Завантаження БД з Drive на Colab
import shutil
shutil.copy('/content/drive/MyDrive/telegram_messages.db', '/content/telegram_messages.db')

conn = sqlite3.connect('/content/telegram_messages.db')

query = '''
    SELECT m.id, m."Джерело", m."Текст повідомлення", src.id as source_id
    FROM messages m
    JOIN sources src ON src.name = m."Джерело"
    WHERE NOT EXISTS (
        SELECT 1 FROM message_strategies ms
        WHERE ms.message_id = m.id AND ms.source_id = src.id
    )
'''

df_unlabeled = pd.read_sql_query(query, conn)
print(f"Нерозмічених повідомлень: {len(df_unlabeled)}")

... Нерозмічених повідомлень: 9031
```

Рисунок 3.3.2. Завантаження і вибірка нерозмічених повідомлень

Передбачення здійснюються групами по 32 повідомлення одночасно. Для кожного повідомлення модель повертає вектор із шести значень від 0 до 1 – по одному на кожну стратегію. Ці значення порівнюються з оптимальними

порогами, визначеними на етапі навчання: інформативна – 0.35, оцінно-впливова – 0.40, емоційно-впливова – 0.55, регулятивно-спонукальна – 0.50, конвенційна – 0.65, РЕКЛАМА – 0.70. Якщо значення перевищує поріг – стратегія вважається присутньою у повідомленні (рис. 3.3.3-3.3.4).

```
def predict_batch(texts):
    encoding = tokenizer(
        texts,
        max_length=256,
        padding=True,
        truncation=True,
        return_tensors='pt')

    input_ids = encoding['input_ids'].to(device)
    attention_mask = encoding['attention_mask'].to(device)

    with torch.no_grad():
        outputs = model(input_ids=input_ids, attention_mask=attention_mask)
        probs = torch.sigmoid(outputs.logits).cpu().numpy()
    return probs
```

Рисунок 3.3.3. Функція *predict_batch*

```
# Розмітка порціями
all_probs = []
texts = df_unlabeled['Текст повідомлення'].tolist()
for i in range(0, len(texts), BATCH_SIZE):
    batch = texts[i:i+BATCH_SIZE]
    probs = predict_batch(batch)
    all_probs.extend(probs)
    if i % 500 == 0:
        print(f"Оброблено: {i}/{len(texts)}")
all_probs = np.array(all_probs)
print("Розмітку завершено")
```

Рисунок 3.3.4. Цикл розмітки з порогоми

При збереженні результатів до бази даних застосовується та сама логіка що і при ручному анотуванні: якщо модель визначила мітку РЕКЛАМА – інші стратегії для цього повідомлення не присвоюються, оскільки рекламні повідомлення є взаємовиключними з новинними стратегіями. Повідомлення, для яких жодна стратегія не перевищила порогового значення, залишаються без розмітки (рис. 3.3.5).

```

cursor = conn.cursor()
cursor.execute("PRAGMA foreign_keys = ON")
inserted = 0
skipped = 0

for idx, row in df_unlabeled.iterrows():
    msg_id = row['id']
    source_id = row['source_id']
    probs = all_probs[idx - df_unlabeled.index[0]]
    # Визначаємо які стратегії присутні за порогоми
    assigned = []
    for strategy_idx, (prob, threshold) in enumerate(zip(probs, THRESHOLDS)):
        if prob > threshold:
            assigned.append(strategy_idx + 1) # strategy_id від 1 до 6
    # Якщо РЕКЛАМА (6) присутня – залишаємо тільки її
    if 6 in assigned:
        assigned = [6]
    # Якщо жодної стратегії не визначено – пропускаємо
    if not assigned:
        skipped += 1
        continue
    for strategy_id in assigned:
        cursor.execute('''
            INSERT OR IGNORE INTO message_strategies (message_id, source_id, strategy_id)
            VALUES (?, ?, ?)
        ''', (msg_id, source_id, strategy_id))
        inserted += 1

conn.commit()
print(f"Збережено міток: {inserted}")
print(f"Повідомлень без стратегій: {skipped}")

```

Рисунок 3.3.5. Блок збереження результатів у БД

За результатами автоматизованої розмітки повного корпусу отримано такий розподіл стратегій (рис. 3.3.6):

```

***
Загальна статистика після автоматичної розмітки:
           name  cnt
Інформативна 9016
Оцінно-впливова 2710
Емоційно-впливова 1471
Регулятивно-спонукальна 1430
Конвенційна 844
РЕКЛАМА 384

```

Рисунок 3.3.6. Результат блоку статистики

Інформативна стратегія є домінантною – 9016 присвоєнь, оцінно-впливова посідає друге місце – 2710, емоційно-впливова – 1471, регулятивно-спонукальна – 1430, конвенційна – 844, РЕКЛАМА – 384.

Загалом автоматизовану розмітку отримали понад 9000 повідомлень корпусу, для яких сумарно присвоєсно 14315 міток стратегій. Лише 3

повідомлення залишились без розмітки – для них модель не визначила жодної стратегії з достатньою впевненістю, що свідчить про коректну роботу системи класифікації і відсутність схильності до надмірного присвоєння міток. Автоматично розмічена база даних *telegram_messages_auto.db* розміщена в додатку 1.

Слід зазначити, що автоматична розмітка має певні обмеження точності, особливо для семантично близьких класів. Зокрема, в процесі вибіркової перевірки результатів виявлено окремі випадки хибного присвоєння оцінно-впливової стратегії повідомленням із нейтральною фактологічною подачею. Такі помилки є закономірним наслідком F1-score 0.60 для цього класу і відображають реальну лінгвістичну складність розмежування семантично близьких стратегій, що було зафіксовано ще на етапі оцінки якості моделі.

3.4. Оцінка ефективності автоматизованого аналізу та інтерпретація результатів

Отримана в результаті автоматизованої класифікації розмітка корпусу дозволяє перейти до змістовного лінгвістичного аналізу – виявлення закономірностей у використанні вербальних стратегій впливу як на рівні всього корпусу, так і в розрізі окремих каналів. Для візуалізації отриманих результатів розроблено допоміжний скрипт мовою Python із використанням бібліотек *matplotlib* та *pandas*, код якого розміщено у Додатку 4.

- **Загальний розподіл стратегій у корпусі**

Загальний розподіл вербальних стратегій у корпусі підтверджує домінування інформативної функції в українському новинному Telegram-дискурсі воєнного часу (рис. 3.4.1).

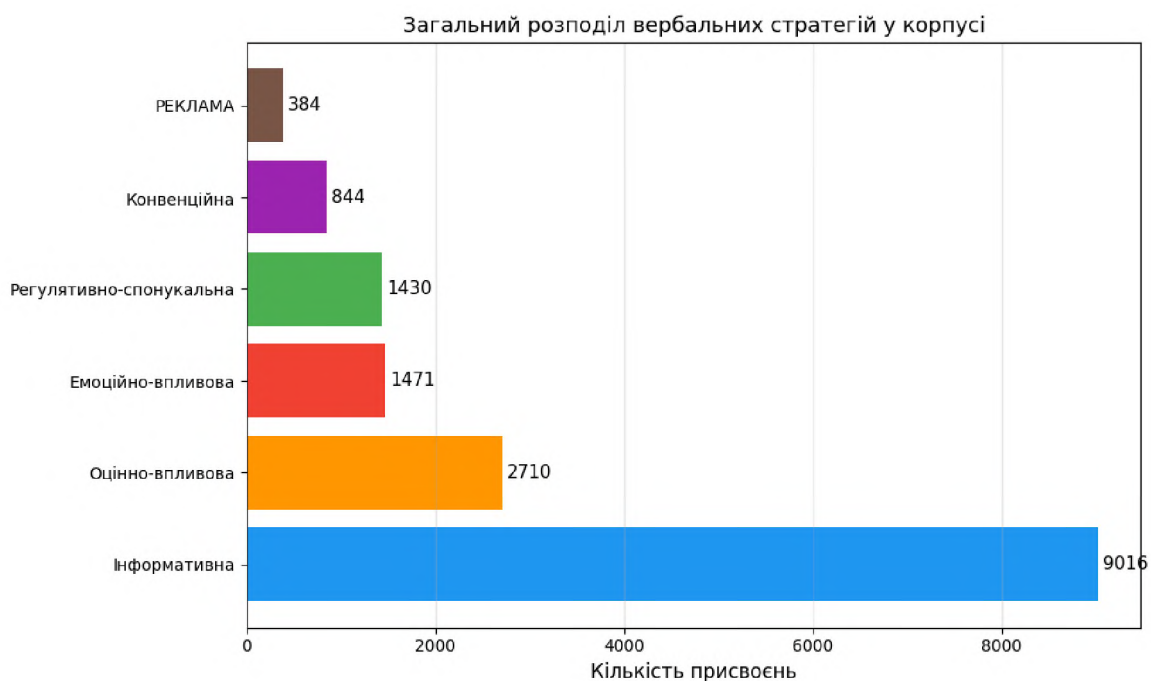


Рисунок 3.4.1. Загальний розподіл вербальних стратегій у корпусі
Telegram-повідомлень

Інформативна стратегія охоплює 9016 присвоєнь – абсолютна більшість серед усіх класів. Це закономірно відображає жанрову специфіку новинного дискурсу, де передача фактологічної інформації є базовою комунікативною функцією. Оцінно-впливова стратегія посідає друге місце з 2710 присвоєннями, що свідчить про високу частотність оцінних суджень в умовах інформаційного протистояння воєнного часу – нейтральна подача поступається місцем ангажованій інтерпретації подій. Емоційно-впливова (1471) та регулятивно-спонукальна (1430) стратегії представлені приблизно рівномірно, що вказує на паралельне використання емоційного тиску і прямих закликів до дій як взаємодоповнювальних інструментів впливу. Конвенційна стратегія (844) і РЕКЛАМА (384) є найрідкіснішими, що повністю відповідає їхній функціональній обмеженості в новинному контексті.

- **Порівняльний аналіз стратегій за каналами**

Порівняльний аналіз розподілу стратегій між п'ятьма каналами виявляє суттєві відмінності, зумовлені типологічними особливостями кожного джерела (рис. 3.4.2-3.4.3).

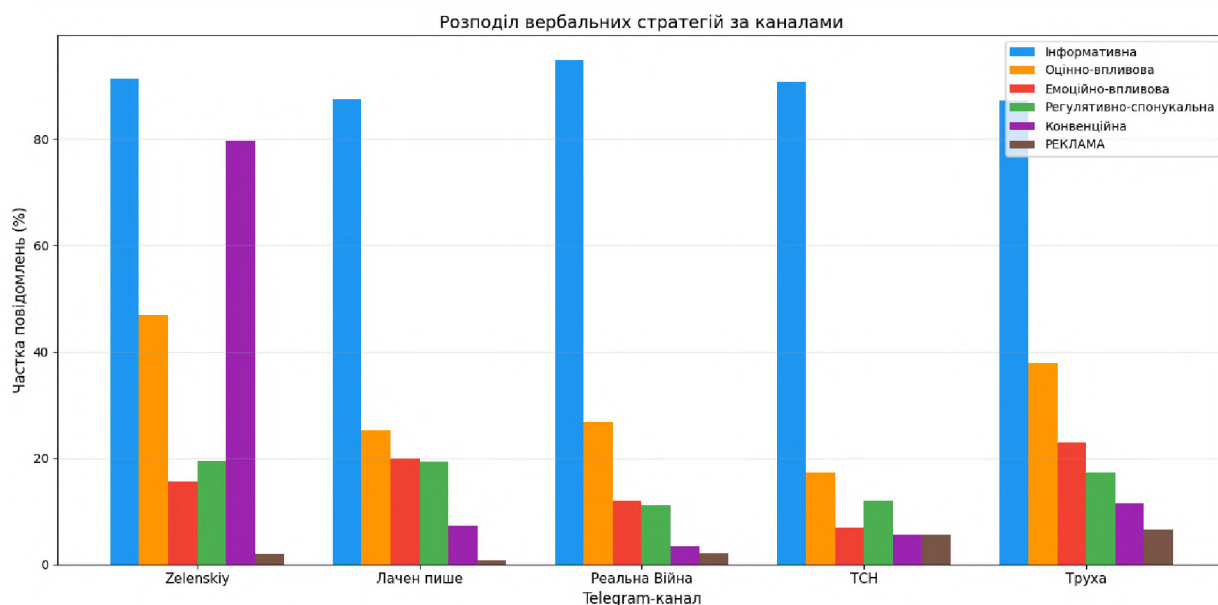


Рисунок 3.4.2. Розподіл вербальних стратегій за каналами (у відсотках від загальної кількості повідомлень каналу)

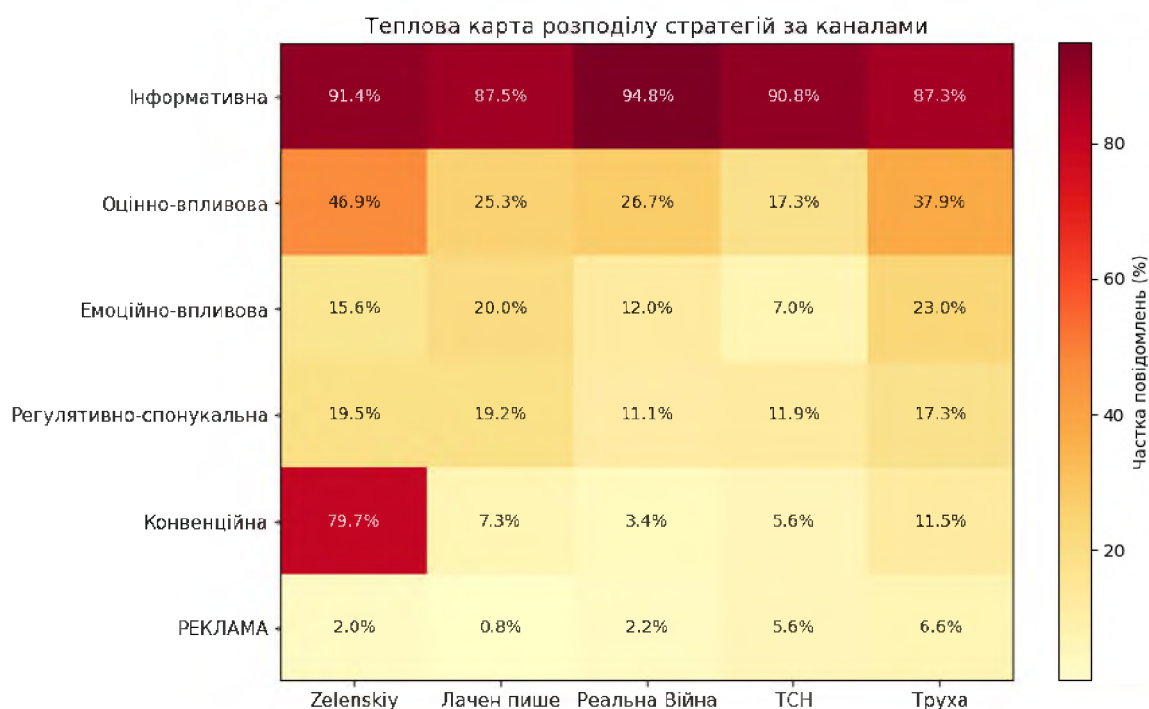


Рисунок 3.4.3. Теплова карта розподілу вербальних стратегій за каналами

Zelenskiy / Official демонструє найвищий серед усіх каналів показник конвенційної стратегії – 79.7%. Це пояснюється жанровою специфікою офіційного президентського каналу: щоденні вечірні звернення Зеленського мають усталену ритуальну структуру з обов'язковими елементами – подяками армії, зверненням до народу, патріотичними формулами на кшталт «Слава

Україні». Такі конструкції є класичними маркерами конвенційної стратегії. Водночас оцінно-впливова стратегія тут також висока – 46.9% – що відображає риторичну природу президентських звернень, спрямованих на формування консолідованого погляду на події війни. Показник РЕКЛАМИ (2.0%) для цього каналу слід розглядати як артефакт класифікації: модель помилково відносила до цієї категорії двомовні повідомлення зі згадками донаційної платформи United24, які за змістом є патріотичними зверненнями, а не рекламою.

Труха ⚡ Україна вирізняється найвищими показниками оцінно-впливової (37.9%) та емоційно-впливової (23.0%) стратегій серед анонімних агрегаторів. Це підтверджує характерну для цього каналу сенсаційну подачу матеріалу з виразною емоційною забарвленістю – саме такий стиль забезпечив каналу широку аудиторію. Водночас Труха має і найвищий серед агрегаторів показник РЕКЛАМИ (6.6%), що відповідає його комерційній моделі монетизації через рекламні публікації.

Лачен пише як авторський канал з верифікованим журналістом демонструє найбільш збалансований розподіл між оцінно-впливовою (25.3%), емоційно-впливовою (20.0%) та регулятивно-спонукальною (19.2%) стратегіями. Це відображає аналітичний характер авторського дискурсу, де факти поєднуються з оцінками, а читач заохочується до певної позиції чи дії. Показово, що конвенційна стратегія тут одна з найнижчих серед усіх, що свідчить про мінімум ритуальних етикетних елементів на користь змістовного контенту.

Реальна Війна | Україна | Новини має найвищий показник інформативної стратегії – 94.8% і найнижчий показник конвенційної – 3.4%. Це цілком відповідає спеціалізації каналу: оперативне документування бойових дій у форматі стислих зведень, карт і фотоматеріалів без риторичних надбудов. Канал функціонує як максимально наближений до нейтрального інформаційного агрегатора в умовах активних бойових дій.

Телеграмна служба новин – Україна демонструє найнижчий показник оцінно-впливової стратегії серед усіх каналів – 17.3% і водночас один із

найвищих показників РЕКЛАМИ – 5.6%. Відносно стримана оцінність при широкому тематичному охопленні (фронт, економіка, політика, культура) вказує на більш нейтральний редакційний стиль порівняно з іншими агрегаторами.

- **Міжканальні закономірності**

Порівняльний аналіз дозволяє виявити кілька системних закономірностей. По-перше, інформативна стратегія є універсальною базою для всіх типів каналів – її частка коливається в межах 87-95% незалежно від типу джерела. По-друге, рівень оцінно-впливової та емоційно-впливової стратегій корелює з типом каналу: найвищий у сенсаційного агрегатора (Труха) і офіційного державного каналу (Zelenskiy), найнижчий у нейтрального агрегатора (ТСН і Реальна Війна). По-третє, конвенційна стратегія є маркером офіційного інституційного мовлення – її різке переважає у каналі Зеленського (79.7%) порівняно з усіма іншими (3.4-11.5%) підтверджує ритуалізований характер президентської комунікації. По-четверте, присутність реклами є виключно характеристикою анонімних агрегаторів – комерційна монетизація через рекламні публікації є нормою для Трухи і ТСН, але практично відсутня у верифікованого автора і офіційного державного каналу.

Таким чином, автоматизована розмітка корпусу засобами машинного навчання дозволила виявити системні закономірності у використанні вербальних стратегій впливу в українському новинному Telegram-середовищі воєнного часу. Отримані результати підтверджують, що тип каналу – анонімний агрегатор, авторський канал чи офіційний державний – суттєво визначає стратегічний профіль його комунікації, що є важливим висновком як для лінгвістичного аналізу медіадискурсу, так і для практичних потреб медіаграмотності та інформаційної безпеки.

Висновки до третього розділу

Третій розділ бакалаврської роботи реалізував повний цикл автоматизованого аналізу вербальних стратегій впливу – від ручного анотування до машинної класифікації та інтерпретації результатів.

У підрозділі 3.1 описано принципи лінгвістичного анотування корпусу на основі типології Т. Толмачової з адаптованою шестикласовою схемою. Розмічена вибірка охопила 908 повідомлень із 1540 присвоєними мітками, а її нерівномірний розподіл відображає реальну частотність стратегій у новинному Telegram-дискурсі воєнного часу.

У підрозділі 3.2 розроблено систему класифікації на основі fine-tuning моделі bert-base-multilingual-cased. Застосування зваженої функції втрат BCEWithLogitsLoss та індивідуального підбору порогів для кожного класу дозволило досягти загального F1-score 0.75. Найвищі показники якості отримали інформативна стратегія (0.91), РЕКЛАМА (0.90) та конвенційна (0.87) – лексично специфічні класи з чіткими маркерами. Нижчі показники оцінно-впливової, емоційно-впливової та регулятивно-спонукальної стратегій (0.60–0.61) зумовлені їхньою семантичною близькістю, що є об'єктивним лінгвістичним обмеженням, а не недоліком моделі.

У підрозділі 3.3 навчена модель застосована для автоматичної розмітки повного корпусу: 14315 міток присвоєно понад 9000 повідомленням, лише 3 залишились без розмітки. Вибіркова перевірка підтвердила загальну коректність розмітки з очікуваними одиничними помилками для семантично близьких класів.

У підрозділі 3.4 здійснено порівняльний аналіз розподілу стратегій між п'ятьма каналами, який виявив системні закономірності: інформативна стратегія є універсальною базою для всіх типів каналів, тоді як рівень оцінності та емоційності корелює з типом джерела. Конвенційна стратегія виявилась маркером офіційного інституційного мовлення, а реклама – характеристикою виключно анонімних комерційних агрегаторів. Отримані результати підтверджують, що тип каналу суттєво визначає його стратегічний профіль, що є значущим висновком як для лінгвістики медіадискурсу, так і для практики медіаграмотності та інформаційної безпеки.

ВИСНОВКИ

У процесі виконання бакалаврської роботи було досягнуто головної мети дослідження – розроблено систему автоматизованого аналізу вербальних стратегій впливу в українському новинному Telegram-середовищі та виявлено закономірності їх використання на матеріалі реальних даних воєнного часу.

У першому розділі було систематизовано теоретичні засади дослідження вербальних стратегій у медіадискурсі. Встановлено, що стратегія мовленнєвого спілкування є свідомо спланованою послідовністю мовленнєвих дій, спрямованою на досягнення конкретної комунікативної мети. За основу класифікації було взято типологію Т. Толмачової – інформативна, оцінно-впливова, емоційно-впливова, регулятивно-спонукальна та конвенційна стратегії. Також було визначено роль Telegram як домінуючої новинної платформи воєнного часу: відсутність редакційного контролю та алгоритмічне формування інформаційного поля суттєво підсилюють ефективність вербальних стратегій впливу, зокрема маніпуляції та емоційного тиску.

У другому розділі було сформовано емпіричну базу дослідження. Для аналізу обрано п'ять типологічно різних україномовних каналів (Труха ⚡ Україна, Реальна Війна, Телеграмна служба новин, Zelenskiy / Official, Лачен пише). За допомогою розробленого Python-скрипту з бібліотекою Telethon було зібрано 9952 повідомлення за грудень 2023 – січень 2024 року. Багатоетапна процедура мовного очищення забезпечила однорідність корпусу, а сформована реляційна база даних SQLite заклала основу для подальшої класифікації.

У третьому розділі було реалізовано повний цикл автоматизованого аналізу. Здійснено ручну розмітку 908 повідомлень із 1540 мітками шести категорій, на основі якої побудовано класифікатор методом fine-tuning моделі bert-base-multilingual-cased. Загальний F1-score склав 0.75: лексично специфічні класи – інформативна (0.91), РЕКЛАМА (0.90) та конвенційна (0.87) – розпізнаються краще за семантично близькі оцінно-впливову, емоційно-впливову та регулятивно-спонукальну (0.60–0.61). За результатами

автоматичної розмітки повного корпусу понад 9000 повідомленням присвоєно 14315 міток.

Порівняльний аналіз виявив системні закономірності: інформативна стратегія є універсальною базою для всіх каналів (87–95%); рівень оцінності та емоційності корелює з типом джерела; конвенційна стратегія є маркером офіційного інституційного мовлення – її переважання у каналі Зеленського (79.7%) відображає ритуалізований характер президентської комунікації; реклама притаманна виключно анонімним комерційним агрегаторам.

Отримані результати підтверджують, що тип каналу суттєво визначає його стратегічний профіль. Розроблена система може бути застосована для моніторингу медіапростору, виявлення маніпулятивних стратегій та підвищення медіаграмотності. Точність розпізнавання семантично близьких класів може бути підвищена шляхом розширення тренувального датасету та залучення додаткових анотаторів, що дозволить зменшити суб'єктивність розмітки і покращити загальну якість класифікації.

СПИСОК ПОСИЛАНЬ

Авраменко, О.М., 2018. *Українська мова (рівень стандарту) : підруч. для 10 кл. закл. загальн. середн. освіти*. Київ: Грамота. с.68-69.

Андрущенко, В.П., Скубашевська, О.С., 2010. Мовні стратегії як чинник освітніх інновацій. *"Гілея: науковий вісник": Збірник наукових праць*. [онлайн] 35, с.133-147. Режим доступу: <<https://enpuir.npu.edu.ua/bitstream/handle/123456789/14234/Andruschenko.pdf;jsessionid=A4F279B0709DF30CBC668A9098F53582?sequence=1>> [Дата звернення: 18 березня 2026].

Анонім, 2005. *Масс-медиа, сектор безопасности и власть. Роль новостных средств массовой информации в контроле и подотчетности сектора безопасности: научное пособие*. Київ: Ін-т журналістики КНУ ім. Тараса Шевченка.

Анонім, 2011. Вплив засобів масової інформації: психологічні особливості. Реферат. *Освіта.UA*. [онлайн] Режим доступу: <<https://osvita.ua/vnz/reports/journalism/22585/>> [Дата звернення: 20 березня 2026].

Анонім, 2018? *Стратегії мовного маніпулювання у виступах політичних лідерів: Шифр «Маніпулювання»* [pdf]. Режим доступу: <<https://philology.lnu.edu.ua/wp-content/uploads/2019/03/Manipuliuvannia.pdf>> [Дата звернення: 18 березня 2026]

Громадське радіо, 2022. *У чому небезпека «інформаційної бульбашки» та як комфортно із неї вийти?* [онлайн] Режим доступу: <<https://hromadske.radio/podcasts/myslennia-bazova-funktsiia/1065085>> [Дата звернення: 19 березня 2026]

Данько-Сліпцова, А., 2023. *Особенности Telegram как интернет-платформы поширения социально значущей информации*. [онлайн] Режим доступу: <<https://ukrinfospace.knukim.edu.ua/article/view/291193/284748>> [Дата звернення: 7 травня 2026]

Комісар, Л.П., н. д. *Дискурс-аналіз як філософсько-методологічна стратегія в сучасних гуманітарних науках.* [pdf] Київ. Режим доступу: <<https://jrn1.nau.edu.ua/index.php/VisnikPK/article/view/9105/11289>> [Дата звернення: 7 травня 2026].

Костирко, Т.М., Корольова, Т.Д. і Жигалкіна, М.С., 2019. *Основи медіаграмотності: навчальний посібник* [pdf]. Миколаїв: НУК. Режим доступу: <<https://rep.nuos.edu.ua/server/api/core/bitstreams/5991ace1-2229-4905-9959-153b3038c782/content>> [Дата звернення: 18 березня 2026]

Крамар, С., 2025. Рейтинг найпопулярніших telegram-каналів в Україні – рейтинг. *Galera.News.* [онлайн] Режим доступу: <<https://galera.news/rejtyng-najpopulyarnishykh-telegram-kanaliv-v-ukrayini-rejtyng-9885/>> [Дата звернення: 7 травня 2026].

Лачен пише, н. д. [Telegram] Режим доступу: <<https://t.me/lachentyt>> [Дата звернення: 7 травня 2026].

Лук'янова, Л., 2010. *Контент-аналіз як психолого-педагогічна процедура.* [онлайн] Умань : ПП Жовтий. Режим доступу: <<https://lib.iitta.gov.ua/7538/1/%D0%9A%D0%BE%D0%BD%D1%82%D0%B5%D0%BD%D1%82-%D0%B0%D0%BD%D0%B0%D0%BB%D1%96%D0%B7.pdf>> [Дата звернення: 7 травня 2026].

Любимый, Я.В., 1993. *Современное массовое сознание; динамика и тенденции развития.* Київ.

Ляпичева, О.Л., 2016. *Посібник до вивчення дисципліни «Основи теорії мовленнєвої комунікації».* Дніпропетровськ: РВВ ДНУ.

Маслова, Т., н. д. *Принципи класифікації комунікативних стратегій.* [pdf] Київ: Національний технічний університет України «КПІ». Режим доступу: <https://www.kamts1.kpi.ua/sites/default/files/files/maslova_pryntsypu.pdf> [Дата звернення: 18 березня 2026]

Матвійчук, Т.П., Житар, І.В., 2024. Мовні засоби впливу на читача у сучасних ЗМІ. *Закарпатські філологічні студії.* [онлайн] 36, с.38-42. Режим доступу:

<<http://zfs-journal.uzhnu.uz.ua/archive/36/8.pdf>> [Дата звернення: 19 березня 2026]

Новицька, О.А., авт.-уклад., Жарікова, Ю.В. та Голоцукова, Ю.О. рец., 2018. *Основи теорії мовної комунікації: Навчальний посібник*. [pdf] Маріуполь: МДУ. Режим доступу: <https://repository.mu.edu.ua/jspui/bitstream/123456789/2756/1/NP_movna_komunikatsia.pdf> [Дата звернення: 18 березня 2026]

Посмітна, В.В., 2019. Комунікативні стратегії та тактики мас-медійного дискурсу в аспекті потенційної загрози деструктивного мовного впливу. *Науковий вісник Міжнародного гуманітарного університету. Сер.: Філологія*. Київ. с.123-126. <https://doi.org/10.32841/2409-1154.2019.39.3.28>

Реальна Війна | Україна | Новини, н. д. [Telegram] Режим доступу: <<https://t.me/voynareal>> [Дата звернення: 7 травня 2026].

Семигіна, Т.В., 2001. Використання дискурс-аналізу в дослідженнях із соціальної роботи. *НАУКОВІ ЗАПИСКИ*. Том 19. Спеціальний випуск. 2. [pdf] Київ: Національний університет «Києво-Могилянська академія», с. 322–325. Режим доступу: <https://www.socosvita.kiev.ua/sites/default/files/Semigina_2001_1.pdf> [Дата звернення: 7 травня 2026].

Семотюк, О., 2015. Аналіз дискурсу vs. контент-аналіз: методично-методологічні відмінності і подібності. Людина. Комп'ютер. Комунікація : збірник наукових праць. [pdf] Львів: Видавництво Львівської політехніки. с.79-82. Режим доступу: <<https://ena.lpnu.ua/handle/ntb/32510>> [Дата звернення: 7 травня 2026].

Тарасенко, Л. Л., 2017. Цифрове середовище як місце здійснення прав інтелектуальної власності. *Україна в умовах реформування правової системи : сучасні реалії та міжнародний досвід : матеріали II Міжнар. наук.-практ. конф.*, [онлайн] Тернопіль : Економічна думка. с.272-275. Режим доступу: <<http://dspace.wunu.edu.ua/bitstream/316497/21432/1/272-275.pdf>> [Дата звернення: 19 березня 2026].

Телеграмна служба новин - Україна, н. д. [Telegram] Режим доступу: <https://t.me/tgsn_ua> [Дата звернення: 7 травня 2026].

Труха Україна, н. д. [Telegram] Режим доступу: <<https://t.me/truehanewsua>> [Дата звернення: 7 травня 2026]

Федотова, А., 2024. 25 найбільших телеграм-каналів в Україні. *Netpeak Journal*. [онлайн] Режим доступу: <<https://netpeak.net/uk/blog/25-naybil-shikh-telegram-kanaliv-v-ukraini/>> [Дата звернення: 7 травня 2026]

Швелідзе, Л.Д., 2021. *Мовні засоби реалізації комунікативних стратегій у дискурсі соціальних мереж (на матеріалі української й англійської мов): Автореферат*. [pdf] Вінниця. Режим доступу: <<https://abstracts.donnu.edu.ua/article/view/9699>> [Дата звернення: 18 березня 2026]

DB Browser for SQLite, н. д. *DB Browser for SQLite*. [online] Available at: <<https://sqlitebrowser.org/>> [Accessed: 8 May 2026].

Google, н. д. *Google Colab*. [online] Available at: <<https://colab.research.google.com/>> [Accessed: 20 May 2026].

Hugging Face, н. д.-a *Hugging Face*. [online] Available at: <<https://huggingface.co/>> [Accessed: 20 May 2026].

Hugging Face, н. д.-b *bert-base-multilingual-cased*. [online] Available at: <<https://huggingface.co/google-bert/bert-base-multilingual-cased>> [Accessed: 20 May 2026].

JetBrains, н. д. *PyCharm: Getting Started*. [online] Available at: <<https://www.jetbrains.com/help/pycharm/getting-started.html>> [Accessed: 8 May 2026].

Lonami, н. д. *Telethon*. [online] Available at: <<https://github.com/lonamiwebs/telethon>> [Accessed: 8 May 2026].

Python Software Foundation, n. d. *Python*. [online] Available at: <<https://www.python.org/>> [Accessed: 8 May 2026].

SQLite Consortium, n. d. *SQLite*. [online] Available at: <<https://sqlite.org/>> [Accessed: 8 May 2026].

Telegram, n. d.-a *MTPROTO Mobile Protocol*. [online] Available at: <<https://core.telegram.org/mtproto>> [Accessed: 8 May 2026].

Telegram, n. d.-b *Telegram*. [online] Available at: <<https://telegram.org/>> [Accessed: 7 May 2026].

Unknown, 2023. Understanding Media Audience Segmentation. *HIVO*, [online] Available at: <<https://hivo.co/blog/understanding-media-audience-segmentation>> [Accessed: 20 March 2026].

Zelenskiy / Official, n. d. [Telegram] Режим доступу: <https://t.me/V_Zelenskiy_official> [Дата звернення: 7 травня 2026].

ДОДАТКИ

Додаток 1.

**GitHub – платформа для розміщення
та спільної розробки програмного коду,
яка використовувалась для доступу до бібліотеки Telethon.**

Режим доступу: <https://github.com/nmanziy/Diploma>

Додаток 2.

**Збір та попередня обробка даних
з Telegram-каналів засобами машинного навчання**

Режим доступу:  machine_learning.ipynb

Додаток 3.

**Автоматичне визначення
комунікативних стратегій у новинних текстах**

Режим доступу:  auto_strategy.ipynb

Додаток 4.

**Візуалізація статистичних результатів
аналізу комунікативних стратегій**

Режим доступу:  visual_statistics.ipynb