

Міністерство освіти і науки України  
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ «КИЄВО-МОГИЛЯНСЬКА АКАДЕМІЯ»  
Кафедра інформатики факультету інформатики



## Магістерська робота

освітній ступінь – магістр

на тему: «Виявлення наративів у соціальних мережах українською та російською мовами»

Виконав: студент 2-го року навчання  
освітньої програми «Інженерія  
програмного забезпечення»,  
спеціальності 121 Інженерія  
програмного забезпечення

Недосека Данило Станіславович

Керівник: Ігнатенко О. П.,  
доктор фізико-математичних наук,  
старший науковий співробітник

Рецензент: Каніщева О. В.

Кваліфікаційна робота захищена  
з оцінкою \_\_\_\_\_

Секретар ЕК \_\_\_\_\_

« \_\_\_\_ » \_\_\_\_\_ 20 \_\_\_\_ р.

Міністерство освіти і науки України  
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ «КИЄВО-МОГИЛЯНСЬКА АКАДЕМІЯ»  
Кафедра інформатики факультету інформатики

ЗАТВЕРДЖУЮ

Зав. кафедри інформатики

к.ф-м.н., доц. Гороховський С.С

\_\_\_\_\_ (підпис)

“ \_\_\_\_\_ ” \_\_\_\_\_ 2023 р.

**ІНДИВІДУАЛЬНЕ ЗАВДАННЯ**

на магістерську роботу

студенту 2 р.н. магістерської програми Інженерія Програмного Забезпечення  
Недосеці Данилу Станіславовичу  
Розробити Систему автоматичного створення анотованих наборів даних для  
виявлення наративів російською та українською мовою на основі  
великих мовних моделей

Зміст текстової частини до магістерської роботи:

Зміст

Анотація

Вступ

1 Огляд літератури

2 Розробка системи створення анотованих наборів даних

3 Проведення експериментів

Висновки

Список використаних джерел

Додатки

Дата видачі “ \_\_\_\_\_ ” \_\_\_\_\_ 2023 р.

Керівник

О.П. Ігнатенко, доктор фіз.-мат. наук, старший науковий співробітник

\_\_\_\_\_ (підпис)

Завдання отримав

Д.С. Недосека

\_\_\_\_\_ (підпис)

## Графік підготовки кваліфікаційної роботи до захисту

Графік узгоджено «\_\_\_\_\_» \_\_\_\_\_ 2023 р.

| № з/п | Перелік робіт  | Термін     | Підпис | Дата | Примітка |
|-------|--|------------|--------|------|----------|
| 1.    | Отримання завдання на дипломну роботу                    | 20.10.2023 |        |      |          |
| 2.    | Огляд теоретичної літератури про великі мовні моделі     | 18.11.2023 |        |      |          |
| 3.    | Огляд існуючих підходів до виявлення наративів           | 04.02.2024 |        |      |          |
| 4.    | Огляд особливостей виявлення наративів українською мовою | 01.02.2024 |        |      |          |
| 5.    | Реалізація підходів до вирішення поставленої задачі      | 28.03.2024 |        |      |          |
| 6.    | Проведення експериментів                                 | 01.04.2024 |        |      |          |
| 7.    | Написання текстової частини роботи                       | 11.05.2024 |        |      |          |
| 8.    | Корегування роботи за результатами попереднього захисту  | 28.05.2024 |        |      |          |
| 9.    | Оформлення роботи та слайдів                             | 30.05.2024 |        |      |          |
| 10.   | Захист кваліфікаційної роботи                            |            |        |      |          |

## ЗМІСТ

|  |    |
|--|----|
| Анотація .....   | 5  |
| Вступ.....   | 6  |
| РОЗДІЛ 1. ОГЛЯД ЛІТЕРАТУРИ.....  | 9  |
| 1.1 Великі мовні моделі (LLM).....                                     | 9  |
| 1.1.1 Визначення.....  | 9  |
| 1.1.2 Архітектура.....   | 9  |
| 1.2 Виявлення наративів.....   | 10 |
| 1.2.1 Визначення і сфера застосування.....                             | 10 |
| 1.2.2 Існуючі підходи до виявлення наративів.....                      | 10 |
| 1.3 Анотування наративів.....  | 25 |
| РОЗДІЛ 2. РОЗРОБКА СИСТЕМИ СТВОРЕННЯ АНОТОВАНИХ НАБОРІВ<br>ДАНИХ ..... | 27 |
| 2.1 Постановка задачі .....  | 27 |
| 2.2 Опис набору даних .....  | 28 |
| 2.3 Опис підходів до вирішення задачі.....                             | 29 |
| 2.3.1 Автоматична анотація.....  | 29 |
| 2.3.2 Генерація синтетичних даних .....                                | 32 |
| 2.4 Опис реалізації запропонованих підходів.....                       | 32 |
| 3. ПРОВЕДЕННЯ ЕКСПЕРИМЕНТІВ.....                                       | 34 |
| 3.1 Хід експериментів.....   | 34 |
| 3.1.1 Автоматична анотація даних.....                                  | 34 |
| 3.1.2 Генерація синтетичних даних .....                                | 35 |
| 3.2 Результати експериментів .....                                     | 36 |
| Висновки.....  | 39 |
| Список використаної літератури .....                                   | 41 |

## Анотація

Робота присвячена дослідженню можливості використання великих мовних моделей для створення анотованих наборів даних для виявлення заданих користувачем наративів українською та російською мовами в соціальних мережах. У ході роботи проаналізовано існуючі рішення для виявлення наративів, специфіку виявлення наративів українською та російською, а також існуючі рішення для автоматичної анотації наративів. Запропоновано підходи для створення розмічених наборів даних на основі автоматичної анотації нерозмічених даних та генерації синтетичних анотованих даних із використанням великих мовних моделей.

Ключові слова: велика мовна модель, синтетичні дані, наратив, класифікація тексту, обробка природної мови, анотація, JSON, LLaMa, Mistral, GPT-4o, Claude 3 Opus

## Вступ

**Актуальність.** Соціальні мережі в реаліях сьогодення відіграють все більшу роль у спілкуванні, розповсюдженні інформації, вираженні й формуванні суспільних і політичних переконань, що робить їх незамінним джерелом даних для досліджень, але водночас і потужним інструментом у руках зловмисників. Автоматизоване виявлення наративів у соціальних мережах здатне допомогти як в аналізі соціальних тенденцій, пов'язаних із різними сферами життя від політики й економіки до медицини, так і в модерації соціальних мереж, пошуку джерел і поширювачів дезінформації, пропаганди, протистоянні інформаційно-психологічним операціям тощо. Останні зазначені сфери застосування особливо важливі для України в умовах протистояння російській агресії. При цьому наративи постійно змінюються, створюються нові фейки, розповсюджуються нові маніпуляції та викривлення фактів на основі реальних новин. За таких умов у рішеннях для виявлення наративів важлива гнучкість, можливість розпізнавання довільних наративів, заданих користувачем, а не обмежених заздалегідь. Однак наразі для такої задачі українською та російською мовами ані готових рішень, ані наборів даних, точно розмічених для тренування або тестування мовних моделей, поки не існує. Однак розвиток як виявлення наративів іншими мовами, так і обробки природної мови українською, вказують на потенціал досліджень даного напрямку.

**Мета дослідження.** Розробити підходи до автоматичного створення розмічених наборів даних для виявлення заданих наративів у соціальних мережах українською та російською мовами з використанням великих мовних моделей, реалізувати прототип системи створення розмічених даних на основі розроблених підходів і створити розмічені набори даних із використанням реалізованої системи.

**Завдання дослідження.** Проаналізувати існуючі підходи до виявлення та анотації наративів, здійснити аналіз особливостей вирішення цих задач для української та російської мов, визначити найкращі підходи, які не потребують анотованих вручну тренувальних даних, обрати спосіб подання задачі виявлення заданих наративів, обрати підходи та інструменти для створення наборів розмічених даних та оцінити результати використання даних підходів є завданнями цього дослідження.

**Об'єкт дослідження.** Системи виявлення наративів із використанням різних підходів, таких як засновані на правилах підходи, традиційне машинне навчання, глибоке навчання і застосування великих мовних моделей, і системи автоматичної анотації текстів для виявлення наративів, застосування таких систем до даних українською та російською мовами є об'єктами дослідження в даній роботі.

**Предмет дослідження.** Підходи до виявлення наративів та автоматичної анотації, а саме використовувані системи класифікації, моделі машинного навчання, схеми анотацій, способи попередньої обробки та подання вхідних даних, формати запитів до великих мовних моделей, а також можливість створення анотованих даних для виявлення заданих наративів у соціальних мережах українською та російською мовами із використанням великих мовних моделей без додаткового тренування є предметом дослідження в цій роботі.

**Джерела дослідження.** Електронні ресурси, такі як наукові статті та книги з доступом в електронному вигляді, програмна документація, блоги великих технічних компаній, спеціальні форуми й віртуальні конференції є джерелами інформації в цій роботі.

**Наукова новизна одержаних результатів** полягає у реалізації автоматичного створення набору даних, розмічених для розпізнавання саме довільних обраних користувачем наративів, а не визначених заздалегідь, для української та російської мов, а також у впровадженні для цього саме таких

методів, що не потребують розмічених вручну даних для тренування, на основі великих мовних моделей.

**Практичне значення одержаних результатів.** Запропоновані підходи можуть використовуватися для створення розмічених наборів даних для тренування і тестування мовних моделей на задачі виявлення наративів, дозволяючи користувачу обирати наративи, актуальні саме для його цілей і даних. Створення таких наборів даних саме для української та російської мов може допомогти знизити дефіцит ресурсів, характерний для цих мов, і допомогти моделям із відкритим вихідним кодом наблизитися до рівня пропрієтарних рішень у роботі з даними мовами. Також запропоновані підходи гнучкі й можуть бути застосовані до інших мов, які підтримують попередньо натреновані великі мовні моделі.

## **РОЗДІЛ 1. ОГЛЯД ЛІТЕРАТУРИ**

### **1.1 Великі мовні моделі (LLM)**

#### **1.1.1 Визначення**

Великі мовні моделі або LLM – це мовні моделі, засновані на нейронних мережах із особливо великою кількістю параметрів (від десятків мільйонів до мільярдів), які пройшли самокероване або напівкероване навчання на великій кількості немаркованого тексту [1].

#### **1.1.2 Архітектура**

Найбільш широко розповсюдженими архітектурами великих мовних моделей наразі є засновані на трансформерах, особливо таких, що складаються лише з декодера.

Трансформер - це представлена компанією Google архітектура глибокого навчання, яка застосовує механізм багатоголової уваги, вперше запропонований у статті 2017 року "Attention Is All You Need" [2]. У трансформерах текст перетворюється на числові представлення, які називаються токенами, і кожен токен перетворюється на вектор шляхом пошуку в таблиці вкладення слів. На кожному рівні кожен токен потім контекстуалізується в межах контекстного вікна з іншими (незамаскованими) токенами із застосуванням паралельного багатоголового механізму уваги. Це дозволяє посилити сигнал для ключових токенів і послабити сигнал для менш значущих токени.

Основними складовими трансформера є токенизатор для перетворення тексту на токени, єдиний шар вкладення для перетворення токенів і позиції токенів у векторні представлення, шари-трансформери, які здійснюють повторні перетворення векторних представлень, витягуючи все більше і більше лінгвістичної інформації, і шар вилучення для перетворення векторних

представлень у розподіл імовірностей над токенами. Вони складаються з шарів уваги та шарів зворотного зв'язку, що чергуються.

Шари-трансформери складаються з шарів уваги та шарів зворотного зв'язку, що чергуються. В оригінальній роботі Google було представлено і використано шари-трансформери двох типів: енкодер і декодер. Однак далі стало розповсюдженим включення лише одного з двох типів. BERT [3] є прикладом моделі лише з енкодером; а GPT [4] – прикладом моделі лише з декодером.

## **1.2 Виявлення наративів**

### **1.2.1 Визначення і сфера застосування**

Виявленням наративів називають категорію задач із ідентифікації та аналізу наративів у текстах, таких як класифікація текстів за ознакою наративності [5], виявлення ступеня наративності тексту [6], виявлення дезінформації та пропаганди [7], виявлення упереджень [8], сумаризацію, висновки на основі здорового глузду, виявлення подій [5] тощо. У контексті даної роботи розглядається виявлення відомого наративу, тобто визначення, чи виражений заданий наратив у тексті або який зі списку заданих наративів у ньому виражений.

### **1.2.2 Існуючі підходи до виявлення наративів**

Задачі виявлення наративів вимагають розуміння складних взаємозв'язків у природній мові, через що найактивніший розвиток їх рішень є відносно нещодавнім феноменом, який розпочався завдяки появі трансформерів, і згодом, як наслідок, великих мовних моделей. Однак подібні задачі почали вирішувати ще до представлення даної архітектури, з використанням підходів, заснованих на правилах, і традиційного машинного навчання, перш ніж перейти до тонкого налаштування попередньо натренованих трансформерів, і далі – до застосування сучасних великих мовних моделей. Кожен із даних етапів розвитку підходів до обробки природньої мови заслуговує уваги, оскільки старі підходи з часом не

зникають, а можуть вдосконалюватися, поєднуватися з новими, ставати основою для створення нових, або навіть залишатися більш ефективними для окремих задач.

### 1.2.2.1 Засновані на правилах підходи

Підходи на основі правил до задач обробки природньої мови спираються на людські знання про структуру мови, логіку та словники термінів. Одними з найбільш широко використовуваних правил є засновані на присутності або відсутності тих чи інших термінів у контекстному вікні. Наприклад, визначення того, чи є термін запереченням чи ні, може бути досягнуто за допомогою перевірки наявності слів, що позначають заперечення, в межах контекстного вікна. [9] Для реалізації таких правил може застосовуватися співставлення рядків тексту і шаблонів, у якості шаблонів можуть використовуватись регулярні вирази. [10]

Однією з галузей, в яких найраніше усвідомили потребу в автоматичній обробці наративів і почали працювати над рішеннями, є медицина. Медичні записи часто приймають форму неструктурованих наративних текстів, і навіть експерт може витратити багато часу, щоб виділити з них конкретну інформацію. При цьому від швидкості отримання цієї інформації може залежати життя пацієнта. Тому однією із задач, для вирішення яких активно досліджували засновані на правилах підходи, є виділення клінічних концепцій. У роботі «Clinical Concept Extraction: a Methodology Review» Суньян Фу та інші дослідники з клініки Мейо розглядають і порівнюють різні підходи до виділення клінічних концепцій. [9] У дослідженні порівнюються публікації за 2009-2019 роки. Застосовані в публікаціях підходи до виділення клінічних концепцій розділено на чотири категорії: засновані на правилах, традиційне машинне навчання, гібридні, глибоке навчання. Для заснованих на правилах підходів у роботі наводиться наступний перелік процесів:

- Збір клінічних текстів
- Виявлення речень і токенізація

- Застосування правил, логіки, ключових слів
- Ручний перегляд і доповнення лексикону, у випадку низької якості результатів повернення до виявлення речень і токенизації
- Результат

За результатами дослідження, засновані на правилах підходи використовувалися більше ніж будь-яка інша окрема категорія протягом усієї декади. Єдиним винятком став 2011 рік, в якому переважали гібридні методи, однак за 2011 рік було розглянуто всього 4 публікації, тоді як за будь-який інший – не менше 9, що могло негативно вплинути на його репрезентативність відносно загальних тенденцій. Але якщо до 2017 року підходи на основі правил та гібридні підходи застосовувалися в понад 60% публікацій, то починаючи з 2017 року почався тренд на більш широке використання глибокого навчання для виділення клінічних концепцій, і частка підходів з використанням правил почала спадати. У 2019 році в половині публікацій використовували підходи на основі правил або гібридні підходи, тоді як інша половина повністю поклядалася на машинне навчання. Із усіх розглянутих підходів, які використовували правила, станом на 2019 рік лише один гібридний підхід досяг найкращих результатів на свій час на одному завданні. У ньому застосовувалися умовне випадкове поле (англійською conditional random field, CRF) та опорно-векторна машина (англійською support vector machine, SVM) у поєднанні з контекстним рушієм. Конкретною задачею, на якій цей підхід показав себе найкращим, було виявлення медикаментів, інструкцій з їх прийняття й поясень до інструкцій. Таким чином, для задач виділення клінічних концепцій засновані на правилах методи переважно втрачають актуальність і здатні були досягати результатів рівня state-of-the-art у 2019 році лише в поєднанні з машинним навчанням.

Іншою задачею обробки наративних медичних текстів є виявлення секцій. Виявлення секцій як у структурованих, так і неструктурованих текстах здатне

полегшити інші пов'язані задачі, такі як виділення клінічних концепцій, дозволяючи шукати інформацію не по всьому тексту, а лише в потрібній секції. У публікації «Current approaches to identify sections within clinical narratives from electronic health records: a systematic review» Александра Помарес-Кімбая, Маркус Кройцталер і Стефан Шульц проаналізували 39 досліджень, спрямованих на вирішення даної задачі [10]. Так само, як у випадку виділення клінічних концепцій, підходи на основі правил були найбільш розповсюджені, становлячи 59% від усіх проаналізованих. У цих підходах використовувалися наступні категорії правил:

- пряме співставлення заголовків секцій із термінами зі словника
- використання регулярних виразів, які відповідають певним типам тексту
- ймовірнісні правила, що використовують такі дані, як Баєсові ймовірності появи слів у кожній із секцій обраного тексту
- ієрархічні правила, які використовують ієрархічні зв'язки між термінами зі словника
- словники синонімів і варіацій слів

Згідно висновкам дослідження, засновані на правилах підходи давали більш якісні результати, ніж гібридні підходи й чисте машинне навчання, однак мали більш вузьку спеціалізацію. Такі результати демонструють, що більш ранні підходи, засновані на правилах, можуть і перевершувати традиційне машинне навчання у специфічних задачах.

Якщо розглянути вище роботи дозволяють порівняти засновані на правилах підходи з традиційним або глибоким машинним навчанням, то у публікації «Extracting Social Support and Social Isolation Information from Clinical Psychiatry Notes: Comparing a Rule-based NLP System and a Large Language Model» Браджа Гопал Патра, Лорен А. Лепов та інші порівнюють їх із великою мовною моделлю Flan-T5-XL[11][12]. За результатами експериментів система на основі правил

показала себе суттєво краще за Flan-T5 на одному з використаних наборів даних (f-score 0.90 у порівнянні з 0.62) і на рівних на другому наборі даних (f-score 0.82 у порівнянні з 0.81). Дослідники інтерпретували такий результат як наслідок того, що розробка лексику для системи на основі правил відбувалася в тандемі з розробкою правил для ручної анотації. Таким чином, навіть у порівнянні з великими мовними моделями засновані на правилах підходи здатні досягати високих результатів за рахунок більшої пристосованості до конкретної задачі.

### **1.2.2.2 Традиційне машинне навчання**

Під традиційним машинним навчанням в контексті обробки природної мови маються на увазі такі підходи, що не застосовують багатошарових нейронних мереж, а отже не можуть вважатися глибоким навчанням. До таких підходів можна віднести використання опорно-векторних машин (англійською support vector machine, SVM), наївних байєсових класифікаторів, стохастичного градієнтного спуску, лінійної або логістичної регресії, k найближчих сусідів тощо.

Опорно-векторні машини — моделі керованого навчання, що використовуються для задач класифікації та регресійного аналізу. Принцип їх роботи полягає в переводі векторів у простір більшої розмірності й пошуку такої гіперплощини, що розділяє гіперплощини векторів із максимальною відстанню до кожної з них. Дану відстань називають зазором, через що SVM називають класифікаторами з максимальним зазором. Їх робота ґрунтується на припущенні, що неперервне збільшення зазору веде до зменшення середньої помилки класифікатора [13].

У статистиці наївні класифікатори Байєса - це сімейство лінійних "імовірнісних класифікаторів", які припускають, що ознаки є умовно незалежними, враховуючи цільовий клас. Сила (наївність) цього припущення і є причиною назви класифікатора. Ці класифікатори є одними з найпростіших моделей байєсових мереж [14].

Лінійна регресія — це статистичний метод, що використовується для моделювання та аналізу відносин між залежною змінною та однією або кількома незалежними змінними. Лінійна регресія виходить з припущення, що існує лінійний зв'язок між змінними. Вона передбачає, що розподіл похибок є нормальним зі середнім значенням, що дорівнює нулю, та гомоскедастичністю, тобто постійною дисперсією залишків [15]. Логістична регресія передбачає результат категоріальної залежної змінної. Вона повинна виражатися категоріальним або дискретним значенням. Логістична регресія розширює лінійну регресію за допомогою "S"-подібної логістичної функції (відомої також як сигмоїдна функція), яка визначає ймовірності. Сигмоїдна функція перетворює будь-яке дійсне число в діапазон між 0 та 1. [16]

У дослідженні методологій виділення клінічних концепцій, яке вже розглядалося в контексті підходів на основі правил [9], також можна побачити тенденції розвитку традиційних підходів машинного навчання для обробки наративів. До 2017 року спостерігалася тенденція поширення використання таких підходів, їх частка серед досліджених сягала 30%, однак із представленням моделі BERT глибоке навчання почало витісняти традиційні методи, їх використання поступово пішло на спад. Однак навіть попри це два підходи із застосуванням традиційного машинного навчання здобули статус найкращих на свій час на двох задачах. Обидва використовували опорно-векторні машини, один із них у поєднанні з підходами на основі правил для розуміння контексту.

Ефективність SVM для задач виявлення наративів демонструють у своєму дослідженні 2019 року також Дірксон та інші.[17] Дана робота фокусувалася на наративах особистого досвіду, пов'язаного з хворобами і лікуванням, якими люди ділилися на форумах, у соцмережах, тощо. Тому перед моделями машинного навчання стояла задача визначення, містить текст такий наратив чи ні. У рамках дослідження для даної задачі бінарної класифікації порівнювали такі моделі як лінійний опорно-векторний класифікатор, мультиноміальний наївний баєсів

класифікатор, стохастичний градієнтний спуск, логістична регресія і  $k$  найближчих сусідів. Усі вони тренувалися на попередньо анотованих експертами даних, показник згоди між експертами становив 0.69. Також порівнювалися різні способи попередньої обробки текстів, такі як лематизація, стемінг, поділ на  $n$ -грами зі значенням  $n$  від 3 до 6, перетворення документів на вектори тощо. Найкращі результати з точки зору показників  $F1$  і повноти показав лінійний опорно-векторний класифікатор на 3-грамах зі значеннями 0.815 і 0.844 відповідно. З точки зору точності найкращим виявився той самий класифікатор, але з використанням 4-грам, досягнувши значення точності 0.844, у порівнянні з 0.793 для 3-грам.

Іншим підходом до розуміння наративів у соціальних мережах є групування подібних за змістом дописів. Одним із способів це зробити є тематичне моделювання – групування текстів за приналежністю до тієї чи іншої теми. У роботі Дірксона для аналізу обговорюваних тем було використано тематичне моделювання з невід'ємною матричною факторизацією та ваговими коефіцієнтами TF-IDF без використання стоп-слів. Темам присвоювалися мітки вручну, шляхом аналізу слів з найбільшою вагою та найбільш релевантних повідомлень у кожній темі. Іншим способом, який може бути задіяний сам по собі або в рамках тематичного моделювання, є ієрархічна кластеризація дописів. Саме його задіяли Соруш Восухі і Деб Рой для розробки напівавтоматичної системи виявлення наративів у соціальних мережах. [18]. Для фільтрації дописів за ознакою вмісту наративів вони скористалися класифікатором категорій мовлення Tweet Acts, який представили раніше. Даний класифікатор використовує логістичну регресію, аби віднести текст до однієї з шести категорій: твердження, рекомендація, вираження, питання, прохання та інше. Наративними вважалися такі тексти, що містять твердження. Для кластеризації наративних дописів було використано метод виявлення перефразувань, який запропонували в 2015 році Аслі Ейджиоглу та Білл Келлер – переможці SemEval-2015 у даній категорії. Їхній метод передбачав

використання SVM для класифікації пар дописів як таких, що є або не є перефразуванням один одного. На основі класифікації всіх можливих пар нарративних текстів за ознакою перефразування будувався неорієнтований граф, ребрами якого виступали зв'язки між перефразуваннями. Далі на основі графу кластери формувалися за методом Лувена для виявлення спільнот. Суть даного методу полягає в жадібній оптимізації модулярності графа [18]. Модулярністю називають міру щільності зв'язків всередині спільноти в порівнянні зі зв'язками між спільнотами. Першим кроком методу є пошук менших спільнот шляхом локальної оптимізації модулярності, а другим – об'єднання вузлів, які утворюють спільноту, в один вузол, та побудову мережі, вузлами якої вже є спільноти. Ці кроки продовжуються, поки не досягнуто максимуму модулярності, після чого ієрархію кластерів-спільнот вважають готовою. Однак аналіз даної ієрархії виконується вручну користувачами, через що система і є лише частково автоматичною. Тим не менш, надання користувачу зручної структурованої репрезентації даних для аналізу нарративів є не менш важливим кроком до їх розуміння, ніж автоматична класифікація, адже дозволяє відкривати нові нарративи, про які користувач не міг знати заздалегідь, аби класифікувати тексти за їх вмістом.

### **1.2.2.3 Підходи на основі глибокого навчання**

Глибоким навчанням називають підмножину машинного навчання, засновану на багатосарових (глибоких) нейронних мережах з навчанням представлень [19]. Застосування глибоких нейронних мереж для виявлення нарративів досягло суттєвих успіхів у порівнянні з класичним машинним навчанням завдяки представленій у 2018 році моделі BERT [3]. Bidirectional Encoder Representations from Transformers або BERT – представлена Google мовна модель на основі архітектури трансформерів. Вона складається з трьох модулів:

- модуль вкладення, який перетворює закодовані токени на масив векторів представлень
- модуль енкодерів, які виконують трансформації над масивом векторів представлень
- модуль вилучення, який перетворює остаточні представлення на закодовані токени

BERT було попередньо натреновано на задачах моделювання мови і передбачення наступного речення. Згодом на її основі було представлено моделі, які мали суттєві переваги над оригінальною версією. Представлена в 2019 році DistilBERT відрізнялася на 40% меншим розміром, зберігаючи 97% здатності розуміння мови й працюючи на 60% швидше [20]. А модель RoBERTa була отримана в результаті дослідження попереднього тренування BERT, яке виявило суттєві недоліки в тренуванні оригінальної моделі, завдяки роботі над якими RoBERTa досягла результатів рівня state-of-the-art на бенчмарках GLUE, RACE і SQUAD [21].

У своєму дослідженні 2022 року Ганті та інші порівняли здатність різних моделей машинного навчання виявляти наративи в публікаціях у соціальних мережах про рак грудей з використанням експертних анотацій як тренувальних даних.[5] DistilBERT, BERT і RoBERTa порівнювалися з класичними моделями, такими як наївний байєс, опорно-векторні машини і логістична регресія. Оскільки визначення наявності або відсутності наративу в тексті є задачею бінарної класифікації, а не моделювання мови, до кожної з моделей на основі BERT було додано шар класифікації, який приймав на вхід репрезентацію токена [CLS], яким моделі позначали клас. Кожна з моделей на основі BERT показала себе краще за всі традиційні моделі з точки зору показників F1 і рівня відгуку, а DistilBERT і RoBERTa також досягли максимальної точності. Однак для даної задачі саме оригінальна BERT показала себе найкраще загалом, продемонструвавши

найвищий показник F1, а саме 0.988, а також максимальний рівень відгуку. Такі результати вказують на суттєві переваги глибокого навчання над традиційними методами для задачі виявлення наративів.

Еффі Леві та інші у своїй роботі представили нову задачу обробки природної мови, а саме виявлення наративних елементів у текстах, і підходи до її вирішення з використанням глибокого навчання. [22] Для виявлення було обрано такі наративні елементи як «ускладнення», «розв'язка» та «успіх» як найбільш характерні для набору даних, який складався зі статей новин. Задачу було представлено двома способами, ефективність яких порівнювалася: як три окремі задачі бінарної класифікації за кожним із елементів або як одну задачу класифікації з багатьма мітками. У рамках дослідження порівнювалися моделі DistilBERT, BERT і RoBERTa. Для тонкого налаштування для класифікації до моделей було додано багатосаровий перцептрон над вихідними представленнями моделей, який мав один прихований шар, а вихідний шар складався з однієї сигмоїди для бінарної класифікації або трьох сигмоїд (по одній для кожної наративної категорії) для класифікації з багатьма мітками. Для оптимізації в ході тренування моделі використовувався алгоритм AdamW та функція бінарної крос-ентропії. За результатами проведених експериментів класифікація з багатьма мітками виявилася ефективнішою за розбиття на окремі бінарні класифікації через кореляції між наративними категоріями. Серед моделей найкращі результати продемонструвала RoBERTa, середні значення F1 і точності якої становили 0.77. Дане дослідження демонструє можливість використання шлибокого навчання не лише для виявлення наративів, а й для аналізу їхнього змісту, розуміння, які саме наративи містить текст. Однак воно також дає змогу побачити різницю в складності між виявленням наративів як таких і конкретних наративних категорій, адже навіть показники найкращої моделі з найкращим підходом до класифікації суттєво нижчі, ніж для задачі виявлення наративів у розглянутому вище дослідженні Ганті.

У 2023 році Марія Антоняк та інші розробили підхід до експертної анотації проміжків тексту, які містять наративи, і порівняли здатність виявляти нарративні проміжки таких моделей машинного навчання як SVM, RoBERTa, GPT-3.5-Turbo [23] і GPT-4 [24]. Якщо SVM тренувалася, а RoBERTa тонко налаштовувалася на експертних анотаціях, то для моделей GPT використовувалися такі підходи як Zero-Shot і Few-Shot [25]. Найкращі результати продемонструвала RoBERTa з показником F1 у 0.86, випереджаючи GPT-4 з підходом Zero-Shot, яка посіла друге місце з F1 у 0.75. Дане дослідження демонструє ефективність моделей на основі BERT, налаштованих під конкретну задачу, для виявлення нарративів, однак також вказує на потенціал і гнучкість більш сучасних великих мовних моделей, враховуючи їхні результати без тонкого налаштування.

#### **1.2.2.4 Застосування LLM для виявлення нарративів**

Широке застосування авторегресивних великих мовних моделей на мільярди параметрів до задач обробки природньої мови почалося з представлення OpenAI моделі GPT-3, яка після попереднього тренування і без тонкого налаштування змогла продемонструвати високі результати на багатьох різних задачах, таких як переклад, відповіді на питання, використання нового слова в реченні, розшифровка слів тощо [26]. Така гнучкість стала поштовхом для створення нових моделей та їх застосування для різноманітних задач обробки природньої мови, в тому числі для задач, пов'язаних із виявленням нарративів, особливо в контексті політичних новин, пропаганди, інформаційних операцій тощо.

Кіт Бургхардт та інші в своїй роботі досліджували використання великих мовних моделей для анотації дописів із соцмереж, які є частиною інформаційних операцій, інформацією про політичні цілі операцій, атаковані та атакуючі країни та опис задіяних нарративів [27]. Для автоматичної анотації вони використовують GPT-3.5 Zero-Shot, а для автоматичного порівняння результатів з експертною анотацією застосовують такі моделі як BART, GPT-3.5, GPT-4. За результатами

дослідження GPT-3.5 успішно розпізнавала атаковану та атакуючу країну з точністю від 0.6 до 0.8 в залежності від метрики, політичну мету з точністю від 0.1 до 0.4, що не настільки точно, однак вище очікуваного, а також опис подій, тобто виражений у дописі наратив, із точністю приблизно від 0.4 до 0.6 за метриками GPT, маючи суттєво нижчі за базову лінію показники лише за визначенням категорії політичної мети. Дане дослідження показало перспективність використання великих мовних моделей для виявлення інформаційних операцій, опису їх наративів та визначення задіяних осіб, однак також продемонструвало виклики, такі як складність автоматичної інтерпретації та оцінки відповідей моделі через складність автоматичного виявлення галюцинацій, сильна залежність точності результатів від побудови запитів у поєднанні з практично необмеженою кількістю способів їх написання, а також залежність відтворюваності результатів від версії моделі та стохастичності (температури) вихідних даних.

У роботі «Decoding News Narratives: A Critical Analysis of Large Language Models in Framing Bias Detection»[28] Пасторіно та інші порівнюють здатність великих мовних моделей виявляти упередження в новинах використовуючи Zero-Shot і Few-Shot підходи. Для експериментів було обрано моделі GPT-4, GPT-3.5-Turbo і Flan-T5. За результатами експериментів, найвищі показники точності та F1 (77% і 0.75 відповідно для текстів, про які експерти досягли згоди) мала GPT-4 із застосуванням Few-Shot Prompting, однак вона мала схильність визначати тексти як упереджені, якщо в них зустрічалися будь-які ознаки емоційної мови, тоді як GPT-3.5 показала стійкість до емоційної мови при виявленні упереджень. Також дослідники зіткнулися з тими ж викликами, що й Бургхардт у розглянутій вище роботі, при використанні закритих моделей від OpenAI, доступних лише по платному API. при цьому компактніша Flan-T5 із відкритим вихідним кодом показала суттєво слабші результати.

Успішним прикладом виявлення наративів з використанням великих мовних моделей є робота Богдана Павлишенка, зосереджена на аналізі дезінформації та неправдивих новин [29]. Для експериментів у даній роботі використовувалася модель Llama-2 від Meta [30]. Особливостями даної моделі є відкритий доступ, а отже можливість локального розгортання та модифікації, у поєднанні з кількістю параметрів та результатами на задачах обробки природньої мови, порівнюваними з платними закритими аналогами. На відміну від розглянутих вище досліджень, у даній роботі застосовувалося тонке налаштування моделі для поставленої задачі. Задача полягала в аналізі тексту на наявність наративів пропаганди або дезінформації у форматі відкритої відповіді природною мовою із розписаними по пунктах аргументами. Через це даний підхід складно оцінити автоматично, він піддається оцінці експертами, як зазначено в самій роботі. З іншого боку, результати експериментів вказали на здатність моделі Llama-2 генерувати структурований текст із використанням формату JSON, що може суттєво полегшити автоматичну інтерпретацію відповідей моделі за більш строгих вимог до структури відповіді, а також на потенціал моделей сімейства LLaMa в умовах локального розгортання.

Оскільки задачі виявлення наративів часто розглядаються як задачі класифікації, а вхідні дані для великих мовних моделей зазвичай мають формат інструкцій або запитань, логічно розглядати задачу класифікації як питання з варіантами відповіді в контексті LLM. Із цього випливає питання, які підходи ефективніше застосовувати для відповіді на питання з варіантами відповіді з використанням великих мовних моделей. Саме на це питання покликана відповіді робота Робінсона та інших 2023 року [31]. У ній порівнюється підхід, у якому модель розглядає кожен із варіантів відповіді окремо, відомий як *cloze prompting*, і такий, в якому питання та всі варіанти відповіді подаються в одному запиті, який автори називають *multi-choice prompting* або MCP. Для порівняння даних підходів використовуються три моделі: GPT-3, Instruct і Codex.

Експерименти проводилися на трьох наборах даних, і на кожному з них підхід MСР продемонстрував суттєве покращення результатів для моделей Codex та Instruct із найбільшим покращенням на наборі даних RACE-m із 67% до 89.2% для Codex і з 66.8% до 89.6% для Instruct. Однак у випадку GPT-3 застосування цього підходу не призвело до покращення результатів, при цьому Instruct і Codex за використання кожного з підходів перевершували GPT-3. Таким чином, дана робота продемонструвала, що різні великі мовні моделі суттєво відрізняються в своїй здатності до пов'язування символів із варіантами відповіді, однак для моделей, які проявляють її в значній мірі, підхід MСР дозволяє отримувати результати рівня state-of-the-art або близького до нього, випереджаючи підхід cloze prompting.

### **1.2.2.5 Виявлення наративів українською та російською мовами**

Для задач виявлення наративів, як і для задач обробки природної мови загалом, характерний дисбаланс у доступі до даних різними мовами. Через це рішення таких задач для української та російської мов, особливо української, не так широко представлені, як рішення для англійської. Однак із розвитком глибокого навчання, великих мовних моделей і популяризації багатомовних задач обробки природної мови з'явилися нові можливості, і як наслідок нові дослідження в цьому напрямку.

У 2023 році Солопова та інші в рамках свого дослідження розробили та порівняли два підходи до виявлення наративів російської пропаганди в новинах і дописах у месенджері Telegram [32]. Набір даних містив новини та дописи українською, російською, румунською, англійською і французькою. Один із запропонованих підходів був заснований на використанні SVM у поєднанні з підібраним вручну набором ознак і ключових слів для класифікації, тоді як другий підхід задіяв BERT для навчання ознакам і отримання вкладень та лінійну модель для класифікації. Другий підхід показав себе краще в усіх експериментах, окрім навчання на новинах і тестування на дописах із Telegram, здобувши у фінальному

експерименті з навчанням на повному тренувальному наборі даних значення F1 у 0.92. Такий результат демонструє ефективність використання глибокого навчання на основі трансформерів для виявлення наративів українською та російською мовами.

Дементьева, Хиленко та Гро у 2024 році розробили підходи на основі міжмовної передачі знань для класифікації текстів українською мовою [33]. Підходи включали використання систем перекладу, великих багатомовних енкодерів, великих мовних моделей, мовних адаптерів і тестувалися на задачах класифікації токсичності, класифікації формальності та висновків на основі природньої мови. Для тренування і тестування використовувалися два набори даних, один перекладений з англійської, а інший – природньою українською. Підхід із використанням перекладу передбачає застосування готового класифікатора для англійської мови у комбінації з системою перекладу, яка щоразу викликається для перекладу тексту українською на англійську. В результаті перевірки на випадково обраних прикладах для класифікації токсичності було обрано систему перекладу Orpus, а для інших двох задач – NLLB. Даний підхід застосовувався лише для природнього набору даних, аби уникнути можливих негативних наслідків подвійного перекладу. У якості великого багатомовного енкодера було обрано XLM-RoBERTa [34], а в якості великої мовної моделі для використання з підходом zero-shot - LLaMa-2. Підхід із тренуванням адаптера передбачав додавання шару-адаптера для англійської мови до багатомовної мовної моделі, тонкого налаштування для конкретної задачі, а потім заміни адаптера на призначений для української. На задачі класифікації токсичності найкраще себе показала тонко налаштована XLM-RoBERTa як на перекладеному, так і на природньому наборі даних, із F1 і точністю 0.77 на природньому датасеті. Із підходів без додаткового тренування вищі показники F1 і точності на природних українських даних, а саме 0.67, здобула LLaMa-2. На задачі класифікації формальності найефективнішим підходом виявилось тренування адаптера із F1 і

точністю 0.71. Серед підходів без тренування найвищі результати отримав підхід на основі системи перекладу з F1 0.50 і точністю 0.56. Тренування адаптера на даній задачі стало єдиним прикладом кращих результатів на природних даних, ніж на перекладених, серед усіх експериментів даного дослідження. Для висновків на основі природної мови найефективнішою себе показала XLM-RoBERTa із точністю та F1 0.82 на перекладених даних і 0.48 та 0.42 відповідно на природних. Серед підходів без тренування комбінація системи перекладу й класифікатора для англійської також випередила LLaMa-2. Із таких результатів можна зробити висновок, що підходи на основі попереднього тренування ефективніші для класифікації текстів українською мовою, однак навіть вони показали слабкі результати на висновках із природної мови, що може говорити про потребу в більшій кількості тренувальних даних, які в дефіциті для української мови. При цьому в залежності від задачі підходи, які не вимагають тренувальних даних, можуть бути досить ефективними. Серед можливих причин низьких результатів LLaMa-2 також варто зазначити особливості саме цієї моделі з точки зору багатомовності, цілком можливо, що інші моделі або новіші покоління LLaMa здатні на вищі результати для мов із дефіцитом ресурсів, у тому числі української.

### **1.3 Анотування наративів**

Анотація даних для задач виявлення наративів надзвичайно важлива для можливості їх автоматичного вирішення, адже розглянуті вище дослідження вказують на переваги підходів, які передбачають використання тренувальних даних, а також на дефіцит подібних ресурсів для окремих мов, у тому числі української та російської. Однак проводити її вручну на великих обсягах даних, потрібних для тренування великих мовних моделей, може бути занадто затратним з точки зору часу, зусиль, коштів тощо. Через це дослідники активно розробляють підходи для автоматичного або частково автоматичного анотування наративів.

Так, у 2020 році Блекберн та інші розробили підходи до виявлення й анування наративів у соціальних мережах, які передбачали поєднання автоматичних методів із людським втручанням [35]. Для виявлення наративів, за якими мала проводитись аотація, було застосоване автоматичне тематичне моделювання. Виділивши вручну наративи на основі аналізу тематичного моделювання, дослідники перейшли до аотації двох наборів даних. У першому дописи з соцмереж пов'язувалися з наративами автоматично на основі вмісту ключових слів. Для другого набору даних команда з восьми експертів проанотували 11000 дописів вручну, вказуючи їхню позицію (підтримує, проти або нейтрально) відносно кожного з 49 наративів. Вони змогли досягти суттєвого показника згоди у 0.64. Для автоматичної аотації на отриманих наборах даних тренували модель BERT-Base Multilingual Cased, використовуючи поділ на дані тренування й тестування 90% на 10%. Модель досягла показників F1 у 0.66 на аотації наративів і 0.81 на аотації позицій відносно наративів, перевершуючи більш класичні моделі машинного навчання, такі як SVM. Такі результати вказують на ефективність глибокого навчання і великих моделей на основі трансформерів для аотації наративів і демонструють перевагу роботи з контекстом у вигляді тексту допису і тексту наративу в порівнянні з використанням лише тексту допису для віднесення його до одного з класів наративів. Водночас, вони вимагали суттєвого людського втручання для отримання тренувальних даних, а також тренування для специфічної задачі, що спонукає до пошуку більш загальних та менш затратних рішень.

Хакімов і Чіма у своєму дослідженні застосовують повністю автоматизований підхід до аотації наративів у соціальних мережах, пов'язаних з війною росії проти України [36]. Вручну вони тільки підібрали ключові слова для фільтрації дописів Twitter, які стосуються даної теми. Процес фільтрації даних передбачав спершу виключення дописів, що не містять ключових слів, а потім відбір на основі відповіді моделі Flan-T5 на питання, чи пов'язаний пост із темою.

У ході попередньої обробки даних для кожного допису було визначено його мову з використанням готового класифікатора на основі RoBERTa, а потім перекладено його на англійську з використанням системи перекладу NLLB. Це дозволило використовувати готові англомовні рішення для подальшої анотації даних інформацією, пов'язаною з наративами. Для анотації за аналізом настрою (позитивний, негативний або нейтральний) було використано готову попередньо натреновану для роботи з текстами з соцмереж модель. Для визначення позиції дописів відносно воєнного конфлікту, війни, росії та України було використано модель BART [37], тонко налаштованого для висновків із природної мови. Отриманий набір даних було опубліковано з результатами анотації та ID дописів, однак без тексту дописів з етичних міркувань. Дана робота демонструє потенціал підходів на основі перекладу і використання готових англомовних рішень для анотації наративів у багатомовних наборах даних. Такі підходи суттєво полегшують автоматизацію процесу анотації, однак для розуміння, наскільки вони точні, необхідне людське втручання або інші достатньо надійні метрики, яких у рамках даної роботи надано не було.

## **РОЗДІЛ 2. РОЗРОБКА СИСТЕМИ СТВОРЕННЯ АНОТОВАНИХ НАБОРІВ ДАНИХ**

### **2.1 Постановка задачі**

Наразі для української та російської мов спостерігається дефіцит анотованих наборів даних, які б дозволяли тренувати мовні моделі для цієї задачі. Анотація даних вручну – тривалий і трудомісткий процес, що сповільнює розвиток у даному напрямку. Однак, враховуючи досягнення великих мовних моделей у задачах обробки природної мови, можна висунути гіпотезу, що з їх використанням можливо створити розмічений набір даних для виявлення наративів українською та російською мовами без тренування на готових розмічених даних. Задача даної роботи полягає в перевірці гіпотези, що застосовуючи великі мовні моделі

можливо створити анотований набір даних для виявлення наративів українською та російською мовами без додаткового тренування на вже розмічених даних. Для перевірки даної гіпотези необхідно розробити підходи до створення розмічених наборів даних для даної задачі з використанням LLM, реалізувати їх та перевірити їх експериментально, аби підтвердити гіпотезу, спростувати її або дійти висновку про необхідність подальших досліджень.

## 2.2 Опис набору даних

Набір нерозмічених даних, використаний для анотації в даній роботі, є одним із наборів даних зі змагання у виявленні дезінформації, проведеного AI HOUSE [38]. Дані початково були призначені учасникам змагання і дозвіл на їх використання в дослідницьких цілях було надано за запитом за умови не публікувати їх. Набір даних, використаний у даній роботі, має вигляд файлу CSV, що містить записи з дописами й коментарями з месенджера Telegram. Для роботи з даними у форматі CSV було використано бібліотеку Pandas для мови програмування Python. З її використанням вдалося визначити, що файл містить N записів, у кожному з яких такі дані про допис або коментар як канал, на якому його було опубліковано, дата публікації, дата останнього редагування та власне зміст. За датами публікації було визначено, що всі дані зібрані за жовтень 2023 року. Набір даних містить X записів українською і Y записів російською.

Оскільки набір даних містить публікації та коментарі за жовтень 2023-го року, необхідно було обрати для виявлення наративи, які активно поширювалися в даній проміжок часу в соціальних мережах українською та російською мовами. Для цього було використано дані з українського телеграм-каналу Центр Протидії Дезінформації та веб-сторінки Propaganda Diary від Vox Ukraine. Обидва джерела містили інформацію про дати розповсюдження наративів російської пропаганди й інформаційно-психологічних операцій. Однак чим більше варіантів відповіді на питання, тим довший і складніший запит до моделі, що може негативно впливати

на якість відповіді, тому необхідно було виділити обмежену кількість наративів зі знайдених для подальшого використання. Завдяки додатковій інформації від Центру Протидії Дезінформації окрім власне наративів і дат розповсюдження, серед наративів можна було виділити такі, що характерні переважно для російських користувачів, і такі, що також розповсюджували українці, тому мало сенс представити обидві дані категорії як варіанти відповіді для моделі в рівній кількості. Із урахуванням зазначених вище причин було обрано наступні чотири наративи:

1. "Україна продає зброю, надану західними партнерами"
2. "Україні перестануть приділяти увагу та фінансову допомогу через конфлікт між Ізраїлем і Палестиною на Близькому Сході"
3. "Проблеми з енергетичною системою України та відключення світла можуть початися навіть без ракетних обстрілів з боку росії"
4. "Мобілізація в Україні здійснюється насильницькими методами та з порушенням закону"

Перші два наративи були особливо актуальні для російської пропаганди в заданий проміжок часу через події на Близькому Сході, тоді як третій і четвертий також активно розповсюджувалися україномовними користувачами.

## **2.3 Опис підходів до вирішення задачі**

Для вирішення задачі створення анотованого набору даних для класифікації текстів за відповідністю конкретним наративам було задіяно два підходи з використанням великих мовних моделей: автоматичну анотацію існуючого набору даних і створення синтетичного набору даних.

### **2.3.1 Автоматична анотація**

Задачу автоматичної анотації немаркованих було розділено на наступні етапи:

- попередня обробка даних
- власне анотація даних
- перевірка якості отриманої анотації

Попередня обробка даних для анотації для виявлення наративів українською та російською мовами в першу чергу передбачає виділення даних саме цими мовами. Для цієї задачі було вирішено використовувати бібліотеку `Lingua Language Detector` у зв'язку з її здатністю розпізнавати широкий вибір мов, у тому числі українську та російську [39]. Оскільки в наборі даних кількість дописів українською становила майже мільйон, а російською – понад 10 мільйонів, і переважна більшість із них ймовірно за все не стосувалася б жодного із обраних наративів, а також з урахуванням обмежень в обчислювальних ресурсах, наступним кроком попередньої обробки було обрано фільтрацію даних за ключовими словами.

Переходячи до самої анотації було вирішено застосувати великі мовні моделі з підходом `zero-shot`. Однак оскільки існуючий набір даних, використаний у даній роботі, не призначений для відкритого доступу, його необхідно було сприймати як конфіденційні дані та ставити відповідні вимоги до системи автоматичної розмітки. Великим мовним моделям, що надаються по API такими компаніями як `OpenAI` або `Anthropic`, не рекомендується надавати конфіденційні дані, адже якщо вони використають діалоги з користувачами для тренування, це може призвести до випадкового потрапляння таких даних до рук третіх осіб. Через це було прийнято рішення використовувати моделі з відкритим вихідним кодом, розгорнуті локально. Таке рішення означало менші витрати коштів у порівнянні з платними API, однак також призвело до обмеження обчислювальних ресурсів. Для можливості локального використання довелося обирати виключно моделі, кількість параметрів яких не перевищує восьми мільярдів, які підтримують

квантизацію до чотирьох біт. Таким чином, у якості кандидатів було обрано такі моделі як Llama-3-8B від Meta [40], Llama-2-7B для порівняння двох поколінь даної моделі, Mistral-7B від Mistral AI [41], MPT-7B від MosaicML [42], а також Vicuna від Lmsys [43].

Задачу виявлення наративів було вирішено представити як задачу класифікації тексту як такого, що містить один із довільних, заданих користувачем наративів. А її у свою чергу було вирішено розглядати як задачу відповіді на питання з кількома варіантами відповіді, аби моделі, призначені для генерації тексту, могли її вирішувати. Слідуючи розглянутому в даній роботі підходу MCP (Multiple Choice Prompting), було розроблено шаблон запиту до моделі, за яким запит містив текст, який необхідно класифікувати, питання, який з наративів текст містить, варіанти відповіді у форматі «<номер варіанту відповіді>. <текст наративу>», а також початок відповіді, який модель має продовжити, у форматі JSON. Подання відповіді у форматі JSON було вирішено використовувати для полегшення автоматичного виокремлення номера правильної відповіді зі згенерованого моделлю тексту виходячи з аналізу існуючих методів використання LLM для розпізнавання наративів. Також було вирішено, що одним із варіантів відповіді завжди має бути «Жоден із зазначених наративів», адже неможливо гарантувати їх присутність у тексті навіть після фільтрації за ключовими словами.

Стосовно задачі анотації також було висунуто гіпотезу, що анотація на основі консенсусу між моделями здатна покращити результати в порівнянні з окремими моделями, й вирішено перевірити її.

Для надійної перевірки якості автоматичної анотації необхідне людське втручання, однак однією з задач даної роботи є його мінімізація, а саме уникнення ручної анотації великих корпусів тексту. Тому було вирішено випадково обирати невелику частину автоматично анотованих записів і проводити їх анотацію вручну,

аби використовувати їх для валідації як окремих моделей, так і консенсусу між ними, за стандартними метриками, такими як F1, точність, відгук тощо.

### **2.3.2 Генерація синтетичних даних**

Генерація синтетичних даних не має такого елемента конфіденційності, як використання закритого від загального доступу набору даних. У зв'язку з цим для неї було вирішено використовувати найсучасніші на момент проведення експериментів великі мовні моделі, доступні по API, а саме Claude 3 Opus від Anthropic [44], а також GPT-4o від OpenAI [45]. Було вирішено задіяти такі підходи до запитів до моделі, як надання моделі персони експерта в написанні правдоподібних текстів для соціальних мереж від імені системи, а також прохання від імені користувача уявити себе в певній ситуації (наприклад “Уяви, що ти українець і користувач соцмережі, який щойно наткнувся на пост про конфлікт між Ізраїлем і ХАМАС”).

### **2.4 Опис реалізації запропонованих підходів**

Програмну реалізацію запропонованих підходів було виконано мовою Python і розділено на наступні модулі:

- Для автоматичної анотації
  - datasets
  - classification
  - annotation
  - consensus
  - validation
- Для генерації синтетичних даних
  - text\_generation

- `dataset_generation`

Модуль `datasets` містить функції для фільтрації наборів даних за мовою і вмістом ключових слів. Для роботи з наборами даних формату CSV у ньомк використовується бібліотека `Pandas`, а для визначення мов — `Lingua Language Detector`.

У модулі `classification` визначено абстрактний базовий клас `NarrativeClassifier`, на основі якого можна в подальшому розширювати систему класами для інших мовних моделей, та клас `CasualLMClassifier`, який наслідує `NarrativeClassifier` і використовує `text generation pipeline` із бібліотеки `HuggingFace Transformers` для класифікації. На основі цього класу створюються класифікатори для моделей, сумісних із `AutoModelForCasualLM` із бібліотеки `Transformers`, із використанням конфігураційних файлів для різних моделей у форматі `JSON`.

Модуль `annotation` застосовує класифікатори з модуля `classification` для анотації даних, поданих у вигляді типу `Dataframe` бібліотеки `Pandas`, а також конфігурацію для анотації у форматі `JSON`.

Модуль `consensus` містить функцію для виділення із наборів даних, анотованих різними моделями, лише таких записів, для яких частина моделей, вища за заданий поріг, досягла згоди.

У модулі `validation` реалізовано випадковий відбір заданої частини даних для валідації, консольний користувацький інтерфейс для анотації відібраних даних вручну, а також функції для власне валідації на основі отриманої коректної ручної анотації.

Модуль `text_generation` містить базовий абстрактний клас `NarrativeTextGenerator`, який дозволяє створювати класи для генерації синтетичних текстів за заданим наративом із використанням різних моделей. Його наслідують класи `OpenAITextGenerator` і `AnthropicTextGenerator`, призначені для генерації з

використанням API відповідних компаній. Моделі та обмеження на кількість генерованих токенів також задаються в конфігураційних файлах у форматі JSON.

У модулі `dataset_generation` описані вище генератори тексту застосовуються для створення розмічених даних за заданими нарративами з використанням бібліотеки `Pandas` і конфігураційних файлів.

Оскільки для фільтрації за ключовими словами, генерації та розмітки даних використовувалися одні й ті самі нарративи, дані про кожен нарратив, необхідні для всіх процесів, зберігалися разом у файлі `narratives.json`. Окремо зберігалися шаблони запитів до моделей для автоматичної анотації, оскільки вони не залежать від використовуваних нарративів.

### **3. ПРОВЕДЕННЯ ЕКСПЕРИМЕНТІВ**

#### **3.1 Хід експериментів**

##### **3.1.1 Автоматична анотація даних**

На етапі попередньої обробки даних дописи-кандидати для кожного нарративу було підібрано на основі пов'язаних із нарративом ключових слів. Для нарративу про продаж зброї було використано ключові слова і фрази «Україна», «Ізраїль», «ХАМАС», «Ближній Схід», «зброя», «продавати», для нарративу про втрату уваги й допомоги світу - «Україна», «Ізраїль», «Ближній Схід», «забути», «забувати», «увага», для нарративу про відключення світла – «Україна», «енергосистема», «електроенергія», «графік», «відключення», «світло», «електрика», для нарративу про мобілізацію – «мобілізація», «могилизація», «повістка», «бусик», «ТЦК», «ВЛК», «ухилянт», «обмежено придатний», «гарматне м'ясо». На основі вмісту даних ключових слів і фраз для експерименту було відібрано 4000 дописів, із них 1973 українською і 2027 російською.

Перед проведенням експериментів із використанням набору даних, обрані моделі було перевірено на простих штучних прикладах на здатність зв'язно відповідати на запити сформульованого вище вигляду у форматі json, що могло стати проблемою з урахуванням низької кількості параметрів, сильної квантизації та індивідуальних особливостей моделей. В результаті перевірки було виявлено, що МРТ не вдавалося дотримуватися очікуваного форматування відповіді, тоді як відповіді Vicuna не були зв'язними й не містили жодного із запропонованих варіантів відповіді. В результаті для анотації було вирішено використовувати Llama-3-8B, Llama-2-7B та Mistral-7B. Із використанням цих трьох моделей було автоматично анотовано 4000 відібраних за ключовими словами дописів.

На основі припущення, що індивідуальні особливості кожної з моделей можуть призвести до вищої якості результатів у випадку консенсусу між ними, із 4000 класифікованих дописів було відібрано такі, які принаймні 2 з 3 моделей класифікували однаково, і отримано 1548 записів.

Для валідації чверть записів, для яких моделі досягли консенсусу, тобто 387, було випадково вибрано і проанотовано вручну. На основі ручної анотації було проведено оцінку як консенсусу, так і кожної окремої моделі.

### **3.1.2 Генерація синтетичних даних**

Спершу для обох моделей запит формулювався якомога простіше і містив лише прохання українською написати допис або коментар під новиною, що висловлював би певний наратив. Однак із Claude такий підхід не спрацював, модель відповіла спростуванням наративу в такому форматі, наче сприйняла його як особисту думку користувача. Тоді для максимальної ефективності запитів до Claude було вирішено скористатися інструментом Prompt generator від Anthropic [[Prompt generator - Anthropic](#)]. Створений в результаті запит модель зрозуміла, однак відмовилася виконувати з етичних міркувань, оскільки зазначені наративи виражали радикальні погляди на чутливі політичні теми. GPT-4o ж без жодних

заперечень або зайвого тексту генерувала дописи, що доносили задану думку. З одного боку, це допомогло дослідженню, але з іншого підіймає питання про дотримання даною моделлю етичних норм і цінностей OpenAI. Спершу відповіді моделі видавалися занадто формальними, тому від імені системи, а не користувача, моделі було задано персону рядком «You are a social media assistant, skilled in writing natural, believable social media texts using casual language and the kind of vocabulary typically found on social media». Після цього повідомлення почали містити характерне для соціальних мереж неформальне мовлення, використання емодзі й хештегів тощо, однак часто містили словосполучення, взяті напрому з запиту, висловлювали думку занадто прямолінійно. У зв'язку з цим до запиту від імені користувача було додано наступний текст: «Напиши його власними словами, постарайся не повторювати слова і словосполучення, які використав я, а перефразувати їх». Це зробило відповіді більш різноманітними та менш прямолінійними.

### 3.2 Результати експериментів

Результатом експериментів із автоматичної анотації існуючого немаркованого набору даних є оцінка таких показників як точність, F1, відгук, кількість правдивих і неправдивих позитивних і негативних міток для консенсусу моделей LLaMa-3, LLaMa-2 і Mistral по кожному з наративів окремо і загалом або в середньому, а також отримані розмічені дані. Порівняння середніх значень представлено у таблиці 1.

| Model     | F1           | Accuracy     | Precision    | Recall       |
|-----------|--------------|--------------|--------------|--------------|
| Consensus | 0.398        | 0.759        | 0.443        | <b>0.548</b> |
| LLaMa-3   | <b>0.413</b> | <b>0.778</b> | <b>0.519</b> | 0.496        |
| LLaMa-2   | 0.190        | 0.666        | 0.439        | 0.394        |
| Mistral   | 0.256        | 0.722        | 0.350        | 0.345        |

Таблиця 1. Порівняння середніх значень F1, правильності, точності й відгуку для кожної з трьох моделей і консенсусу між ними. Найвищі показники виділено **жирним**.

В середньому LLaMa-3 виявилася найкращою за показниками F1 і точності, демонструючи вищі результати навіть у порівнянні з консенсусом між моделями. Таким чином, гіпотезу про покращення цих показників шляхом узгодження відповідей кількох моделей було спростовано принаймні для випадку, коли одна або деякі з моделей суттєво перевершують інші. Однак шляхом консенсусу все ж вдалось покращити один із показників, а саме відгук. Однак навіть найвище отримане середнє значення F1 менше за 0.5, що свідчить про неготовність протестованих великих мовних моделей до використання для автоматичного виявлення заданих наративів за обмежень локального розгортання й без тонкого налаштування для конкретної задачі. Висока частка правильних міток за результатами класифікації (не нижче 0.666) не є репрезентативною і пов'язана з великою кількістю правильних негативних міток, тобто свідчить лише про те, що при виборі одного з п'яти класів, навіть якщо він невірний, три з чотирьох рішень про те, до якого класу текст не належить будуть вірними. Таке переважання правильних негативних результатів демонструє таблиця 2.

| <b>Model</b> | <b>True positive</b> | <b>True negative</b> | <b>False Positive</b> | <b>False negative</b> |
|--------------|----------------------|----------------------|-----------------------|-----------------------|
| Consensus    | 154                  | 1315                 | 233                   | 233                   |
| LLaMa-3      | 172                  | 1333                 | 215                   | 215                   |
| LLaMa-2      | 64                   | 1225                 | 323                   | 323                   |
| Mistral      | 118                  | 1280                 | 269                   | 269                   |

Таблиця 2. Кількість правдивих і хибних позитивних і негативних рішень для кожної розглянутої моделі

Із показників F1 для кожного окремого наративу, представлених у таблиці 3, також можна зробити висновок, що без додаткового тонкого налаштування використані великі мовні моделі краще справляються з виявленням наративів російською, ніж українською, оскільки результати для наративів, характерних для російськомовних джерел, суттєво вищі.

| Model     | Narrative (short description) |                    |              |                      |       |
|-----------|-------------------------------|--------------------|--------------|----------------------|-------|
|           | Selling weapons               | Support withdrawal | Power surges | Illicit mobilization | None  |
| Consensus | 0.641                         | 0.493              | 0.217        | 0.215                | 0.432 |
| LLaMa-3   | 0.548                         | 0.464              | 0.204        | 0.333                | 0.518 |
| LLaMa-2   | 0.239                         | 0.222              | 0.203        | 0.163                | 0.121 |
| Mistral   | 0.369                         | 0.499              | 0.0          | 0.062                | 0.352 |

Таблиця 3. Значення F1 по кожному окремому наративу для кожної з моделей

Також дані в таблиці 3 розкривають особливості окремих моделей, не очевидні із середніх значень. Так, LLaMa-3 демонструє суттєво кращі результати на російській у порівнянні з українською, тоді як у LLaMa-2 різниця не настільки сильна. У порівнянні з попереднім поколінням, LLaMa-3 має суттєву перевагу в розпізнаванні наративів саме російською, тоді як для української різниця менш суттєва. Також модна побачити, що Mistral суттєво краще справляється з виявленням наративів російською в порівнянні з LLaMa-2, однак абсолютно не придатна для вирішення цієї задачі українською без додаткового тренування. Ще однією особливістю LLaMa-3, яка допомогла їй досягти найвищих результатів, виявилася найвища серед розглянутих моделей здатність виявляти, що в тексті не виражений жоден із запропонованих наративів. З точки зору отриманих анотацій, результатом експерименту є 4000 дописів із соцмереж, проанотованих неточно трьома різними моделями із них 1548 також анотовано за консенсусом між ними, із яких 387 також анотовано вручну.

Результатом експериментів зі створення синтетичних даних став успішно створений набір даних на 5600 записів, по 700 на кожен наратив українською та російською мовами.

## **Висновки**

У результаті виконаної роботи було розроблено два підходи до автоматичного створення анотованого набору даних для виявлення наративів українською та російською мовою. Один із підходів заснований на використанні локально розгорнутих великих мовних моделей для анотації нерозмічених даних, а другий – на генерації синтетичних і одразу розмічених даних з використанням великих мовних моделей, доступних по API.

Підхід на основі автоматичної анотації готових даних на прикладі моделей сімейства LLaMa і Mistral продемонстрував неготовність моделей з відкритим кодом із суттєвими обмеженнями обчислювальних ресурсів до точної автоматичної анотації без додаткового тренування. Результати експериментів також показали, що такі моделі наразі краще пристосовані до виявлення наративів російською мовою, аніж українською, що додатково підкреслило потребу в створенні якісних тренувальних даних для великих мовних моделей українською мовою. Однак навіть отримані внаслідок використання такого підходу неточні анотації можуть бути корисними для навчання мовних моделей. Також окрім неточних автоматично анотованих даних як побічний результат оцінки ефективності моделей було отримано анотовані вручну дані.

Підхід на основі створення синтетичних даних на прикладі GPT-4o продемонстрував, що наразі найбільш дієвим способом створення анотованих даних для виявлення наративів українською та російською є саме синтетична генерація анотованих даних. Головною перевагою такого підходу є максимальна точність анотації, адже анотація прив'язана до запитів до моделі, тож практична

користь отриманих даних залежить саме від якості запитів до моделі, того, наскільки вони сприяють правдоподібності та різноманіттю даних.

Головними викликами в ході роботи стала інтерпретація відповідей великих мовних моделей як міток у задачі класифікації, через що довелося відкинути окремі моделі, обмеження в обчислювальних ресурсах і в кількості можливих запитів до API, пов'язані з обмеженими коштами, а також етичні обмеження великих мовних моделей. Застосування великих мовних моделей по API для створення синтетичних даних має великий потенціал, однак його в будь-який момент може бути заблоковано посиленням фільтрації відповідей таких моделей з етичних причин. З іншого боку, відсутність таких обмежень може надавати нові можливості зловмисникам. Такий конфлікт між інтересами дослідників та компаній, що надають доступ до великих мовних моделей по API, піднімає питання про особливий доступ до таких моделей для верифікованих дослідників, які публікують результати їх використання, із обмеженнями, відмінними від звичайних.

Існують різні шляхи для подальшого розвитку автоматичного створення анотованих даних для виявлення наративів українською та російською. Серед найбільш очевидних тривалі експерименти з текстом запитів до великих мовних моделей для досягнення як кращої анотації існуючих даних, так і кращої генерації синтетичних. Порівняння більшої кількості різних моделей, різних варіантів розміру і квантизації може допомогти виявити моделі, більш придатні для вирішення цієї задачі. Також тонке налаштування LLM з використанням отриманих розмічених даних може допомогти досягти вищих результатів автоматичної анотації. Іншим варіантом є зміна підходу до подання задачі виявлення наративів. У даній роботі її було представлено як задачу класифікації з багатьма класами, однак її також можливо представити як класифікацію з багатьма мітками. Такий підхід може дозволити покращити результати класифікації в

умовах обмежених обчислювальних ресурсів, обробляючи в кожному запиті до моделі лише один із можливих варіантів нарративів, а не всі. Із таким підходом можливо буде також одразу класифікувати текст не просто за наявністю в ньому нарративу, а за позицією автора відносно нарративу.

## Список використаної літератури

1. Better Language Models and Their Implications | OpenAI [Електронний ресурс] <https://openai.com/index/better-language-models/> (дата звернення – 12.05.2024)
2. Attention IS All You Need (arxiv.org) [Електронний ресурс] <https://arxiv.org/abs/1706.03762> (дата звернення – 12.05.2024)
3. BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding (arxiv.org) [Електронний ресурс] <https://arxiv.org/abs/1810.04805> (дата звернення – 14.05.2024)
4. Improving Language Understanding by Generative Pre-Training [Електронний ресурс] [language\\_understanding\\_paper.pdf](#) (дата звернення – 14.05.2024)
5. Narrative Detection and Feature Analysis in Online Health Communities - ACL Anthology [Електронний ресурс] <https://aclanthology.org/2022.wnu-1.7/> (дата звернення – 14.05.2024)
6. [Електронний ресурс] [Computational Detection of Narrativity: A Comparison Using Textual Features and Reader Response - ACL Anthology](#) (дата звернення – 14.05.2024)
7. [Електронний ресурс] [2309.04704 \(arxiv.org\)](#) (дата звернення – 17.05.2024)
8. [Електронний ресурс] [2402.11621 \(arxiv.org\)](#) (дата звернення – 17.05.2024)
9. [Електронний ресурс] [Microsoft Word - manuscript\\_final.docx \(arxiv.org\)](#) (дата звернення – 22.05.2024)
10. [Електронний ресурс] [Current approaches to identify sections within clinical narratives from electronic health records: a systematic review | BMC Medical Research Methodology | Full Text \(biomedcentral.com\)](#) (дата звернення – 22.05.2024)
11. [Електронний ресурс] [\[2210.11416\] Scaling Instruction-Finetuned Language Models \(arxiv.org\)](#) (дата звернення – 22.05.2024)
12. [Електронний ресурс] [2403.17199 \(arxiv.org\)](#) (дата звернення – 22.05.2024)
13. [Електронний ресурс] <https://www.semanticscholar.org/paper/Support-vector-networks-Cortes-Vapnik/24e6cf0796237f21c780a3f0c996817f57b3a1bd> (дата звернення – 22.05.2024)

- 14.[Электронный ресурс] McCallum, Andrew. "[Graphical Models, Lecture2: Bayesian Network Representation](#)" (PDF). (дата звернення – 22.05.2024)
- 15.[Электронный ресурс] [David A. Freedman](#) (2009). Statistical Models: Theory and Practice. [Cambridge University Press](#). p. 26. (дата звернення – 22.05.2024)
- 16.[Электронный ресурс] Tolles, Juliana; Meurer, William J (2016). "Logistic Regression Relating Patient Characteristics to Outcomes". JAMA. **316** (5): 533–4. doi:10.1001/jama.2016.7653. ISSN 0098-7484. OCLC 6823603312. PMID 27483067 (дата звернення – 26.05.2024)
- 17.[Электронный ресурс] <https://ceur-ws.org/Vol-2342/paper3.pdf> (дата звернення – 26.05.2024)
- 18.[Электронный ресурс] [View of A Semi-Automatic Method for Efficient Detection of Stories on Social Media \(aaai.org\)](#) (дата звернення – 26.05.2024)
- 19.[Электронный ресурс] Schulz, Hannes; Behnke, Sven (1 November 2012). "Deep Learning". KI - Künstliche Intelligenz. **26** (4): 357–363. doi:10.1007/s13218-012-0198-z. ISSN 1610-1987. S2CID 220523562 (дата звернення – 26.05.2024)
- 20.[Электронный ресурс] [1910.01108] [DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter \(arxiv.org\)](#) (дата звернення – 26.05.2024)
- 21.[Электронный ресурс] [1907.11692] [RoBERTa: A Robustly Optimized BERT Pretraining Approach \(arxiv.org\)](#) (дата звернення – 26.05.2024)
- 22.[Электронный ресурс] [Detecting Narrative Elements in Informational Text - ACL Anthology](#) (дата звернення – 26.05.2024)
- 23.[Электронный ресурс] [2303.10420] [A Comprehensive Capability Analysis of GPT-3 and GPT-3.5 Series Models \(arxiv.org\)](#) (дата звернення – 26.05.2024)
- 24.[Электронный ресурс] [PDF] [Where Do People Tell Stories Online? Story Detection Across Online Communities | Semantic Scholar](#) (дата звернення – 26.05.2024)
- 25.[Электронный ресурс] OpenAI. 2023. Gpt-4 technical report [2303.08774] [GPT-4 Technical Report \(arxiv.org\)](#) (дата звернення – 26.05.2024)
26. [Электронный ресурс] [PDF] [Language Models are Few-Shot Learners | Semantic Scholar](#) (дата звернення – 26.05.2024)
27. [Электронный ресурс] [2405.03688] [Large Language Models Reveal Information Operation Goals, Tactics, and Narrative Frames \(arxiv.org\)](#) (дата звернення – 26.05.2024)
- 28.[Электронный ресурс] [2402.11621 \(arxiv.org\)](#) (дата звернення – 26.05.2024)
29. [Электронный ресурс] [2309.04704 \(arxiv.org\)](#) (дата звернення – 26.05.2024)
- 30.[Электронный ресурс] [2307.09288] [Llama 2: Open Foundation and Fine-Tuned Chat Models \(arxiv.org\)](#) (дата звернення – 26.05.2024)

31. [Електронний ресурс] [\[PDF\] Leveraging Large Language Models for Multiple Choice Question Answering | Semantic Scholar](#) (дата звернення – 26.05.2024)
32. [Електронний ресурс] [\[2301.10604\] Automated multilingual detection of Pro-Kremlin propaganda in newspapers and Telegram posts \(arxiv.org\)](#) (дата звернення – 26.05.2024)
33. [Електронний ресурс] [\[2404.02043\] Ukrainian Texts Classification: Exploration of Cross-lingual Knowledge Transfer Approaches \(arxiv.org\)](#) (дата звернення – 26.05.2024)
34. [Електронний ресурс] [Unsupervised Cross-lingual Representation Learning at Scale - ACL Anthology](#) (дата звернення – 26.05.2024)
35. [Електронний ресурс] [Detecting and Annotating Narratives in Social Media: A Vision Paper | Semantic Scholar](#) (дата звернення – 26.05.2024)
36. [Електронний ресурс] [\[PDF\] Unveiling Global Narratives: A Multilingual Twitter Dataset of News Media on the Russo-Ukrainian Conflict | Semantic Scholar](#) (дата звернення – 26.05.2024)
37. [Електронний ресурс] [\[1910.13461\] BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension \(arxiv.org\)](#) (дата звернення – 26.05.2024)
38. [Електронний ресурс] [Disinformation Detection Challenge \(aihouse.org.ua\)](#) (дата звернення – 26.05.2024)
39. [Електронний ресурс] [lingua-language-detector · PyPI](#) (дата звернення – 26.05.2024)
40. [Електронний ресурс] [Introducing Meta Llama 3: The most capable openly available LLM to date](#)
41. [Електронний ресурс] [\[2310.06825\] Mistral 7B \(arxiv.org\)](#) (дата звернення – 26.05.2024)
42. [Електронний ресурс] [Introducing MPT-7B: A New Standard for Open-Source, Commercially Usable LLMs | Databricks Blog](#) (дата звернення – 26.05.2024)
43. [Електронний ресурс] [Vicuna: An Open-Source Chatbot Impressing GPT-4 with 90%\\* ChatGPT Quality | LMSYS Org](#) (дата звернення – 26.05.2024)
44. [\[PDF\] The Claude 3 Model Family: Opus, Sonnet, Haiku | Semantic Scholar](#)  
[Електронний ресурс] [\[PDF\] The Claude 3 Model Family: Opus, Sonnet, Haiku | Semantic Scholar](#)
45. [Hello GPT-4o | OpenAI](#) [Електронний ресурс] [Hello GPT-4o | OpenAI](#) (дата звернення – 26.05.2024)