

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук
(повна назва)

Кафедра Штучного інтелекту
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА Пояснювальна записка

рівень вищої освіти другий (магістерський)

Дослідження методів перифразування коротких повідомлень
для соціальних мереж
(тема)

Виконав:
студент 2 курсу, групи СШМ-22-2
Свірчков В.В.
(прізвище, ініціали)

Спеціальність 122 Комп'ютерні науки
(код і повна назва спеціальності)

Тип програми освітньо-наукова
(освітньо-професійна або освітньо-наукова)

Освітня програма Системи штучного інтелекту
(повна назва спеціалізації)

Керівник доц. Турута О.П.
(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри _____
(підпис)

В.О. Філатов
(прізвище, ініціали)

2024 р.

Харківський національний університет радіоелектроніки

Факультет _____ Комп'ютерних наук _____
(повна назва)
Кафедра _____ Штучного інтелекту _____
(повна назва)
Рівень вищої освіти _____ другий (магістерський) _____
Спеціальність _____ 122 Комп'ютерні науки _____
(код і повна назва)
Тип програми _____ освітньо-наукова _____
(освітньо-професійна або освітньо-наукова)
Освітня програма _____ Системи штучного інтелекту _____
(повна назва)

ЗАТВЕРДЖУЮ:
Зав. кафедри _____
(підпис)
« _____ » _____ 20 ____ р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

студентові _____ Свірчкову Володимирі Вікторовичу _____
(прізвище, ім'я, по батькові)

1. Тема роботи _____ Дослідження методів перефразування коротких повідомлень для соціальних мереж _____

затверджена наказом університету від 1 квітня 20 24 р. № 260Ст

2. Термін подання студентом роботи до екзаменаційної комісії 7 червня 20 24 р.

3. Вихідні дані до роботи науково-технічні публікації, дані Інтернет-джерел та відомих наукових проєктів, Python documentation, набір даних для тестування системи _____

4. Перелік питань, що потрібно опрацювати в роботі _____

1) Аналіз предметної галузі _____

2) Опис проведених теоретичних досліджень _____


3) Дослідження наявних методів перефразування _____

4) Опис проведення експериментальних досліджень _____

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
1	Отримання завдання на кваліфікаційну роботу	01.04.2024	виконано
2	Аналіз предметної галузі	05.04.2024	виконано
3	Огляд існуючих методів перефразування	15.04.2024	виконано
4	Проведення експериментального дослідження	10.05.2024	виконано
5	Порівняння наявних методів перефразування	15.05.2024	виконано
6	Написання пояснювальної записки	30.05.2024	виконано
7	Перевірка на академічний плагіат	02.06.2024	виконано
8	Нормоконтроль	03.06.2024	виконано
9	Підготовка презентації та доповіді	04.06.2024	виконано
10	Попередній захист	05.06.2024	виконано
11	Рецензування	06.06.2024	виконано
12	Захист перед ЕК	07.06.2024	

Дата видачі завдання 1 квітня 2024 р.

Студент 
(підпис)

Керівник роботи _____ доц. Турута О.П.
(підпис) (посада, прізвище, ініціали)

РЕФЕРАТ

Пояснювальна записка: 64 с., 6 рис., 16 табл., 2 дод., 25 джерел.

ПЕРЕФРАЗУВАННЯ, КОРОТКІ ПОВІДОМЛЕННЯ, СОЦІАЛЬНІ МЕРЕЖІ, АНАЛІЗ ТЕКСТУ, ШТУЧНИЙ ІНТЕЛЕКТ, АЛГОРИТМИ, ГЛИБИННЕ НАВЧАННЯ, ТРАНСФОРМЕРИ.

Об'єкт дослідження – аналіз методів для перефразування повідомлень у соціальних мережах.

Предмет дослідження – застосування існуючих алгоритмів штучного інтелекту для ефективного перефразування тексту.

Мета роботи – дослідити та оцінити методи перефразування тексту в контексті соціальних мереж, виявити найбільш продуктивні підходи.

Методи дослідження включають аналіз наукової літератури у сфері обробки природної мови, проведення експериментів із застосуванням різних алгоритмів перефразування, порівняння ефективності отриманих моделей.

У ході дослідження було проведено глибокий аналіз існуючих технологій та методів перефразування тексту, з акцентом на обробку коротких повідомлень. Розглянуто основні виклики, пов'язані з точністю та збереженням семантичного змісту при перефразуванні.

Експериментальна частина дослідження демонструє практичне використання різних моделей машинного навчання та алгоритмів штучного інтелекту для задачі перефразування. Особлива увага приділена перекладу, зміні стильового забарвлення та виправленню помилок.

На підставі аналізу результатів експериментів сформульовані рекомендації щодо оптимізації процесу перефразування текстів у соціальних мережах, що сприятиме підвищенню якості комунікації між користувачами.

ABSTRACT

Explanatory note: 64 p., 6 fig., 16 tab., 2 add., 25 references.

PARAPHRASING, SHORT MESSAGES, SOCIAL NETWORKS, TEXT ANALYSIS, ARTIFICIAL INTELLIGENCE, ALGORITHMS, DEEP LEARNING, TRANSFORMERS.

The object of research is to analyze methods for paraphrasing messages on social networks.

The subject of the study is the application of existing artificial intelligence algorithms for effective text paraphrasing.

The purpose of the study is to investigate and evaluate methods of paraphrasing text in the context of social media, to identify the most productive approaches.

Research methods include analyzing scientific literature in the field of natural language processing, conducting experiments using different paraphrasing algorithms, and comparing the effectiveness of the resulting models.

The study conducted an in-depth analysis of existing technologies and methods of text paraphrasing, with a focus on processing short messages. The main challenges related to the accuracy and preservation of semantic content during paraphrasing are considered.

The experimental part of the study demonstrates the practical use of various machine learning models and artificial intelligence algorithms for the paraphrasing task. Particular attention is paid to translation, stylistic changes, and error correction.

Based on the analysis of the experimental results, recommendations are formulated to optimize the process of paraphrasing texts on social media, which will help improve the quality of communication between users.

ЗМІСТ

Перелік умовних позначень, символів, одиниць, скорочень і термінів	7
Вступ.....	8
1 Аналіз предметної галузі	10
1.1 Актуальність соціальних мереж	10
1.2 Огляд існуючих методів перефразування.....	15
1.3 Технології машинного навчання у перефразуванні	18
1.4 Виклики та обмеження методів перефразування.....	21
1.5 Огляд існуючих перефразовувачів	23
1.6 Постановка задачі.....	24
2 Дослідження методів перефразування.....	25
2.1 Трансформери.....	26
2.1.1 BERT.....	28
2.1.2 GPT	32
2.1.3 T5	37
2.1.4 RoBERTa	41
2.2 TextBlob.....	44
2.3 DeepL	46
2.4 Paraphrasing tools	49
2.5 Використані інструменти	51
2.5.1 Transformers	51
2.5.2 PyTorch	51
2.6 Результати роботи.....	52
Висновки	56
Перелік джерел посилання	58
Додаток А Код перефразування повідомлень	62
Додаток Б Відомість кваліфікаційної роботи.....	64

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

ШІ – штучний інтелект;

API – Application Programming Interface – інтерфейс програмування додатків;

BERT – Bidirectional Encoder Representations from Transformers – двонаправлені кодування з трансформерів;

CNN – Convolutional Neural Network – конволюційна нейронна мережа;

DL – Deep Learning – глибоке навчання;

GPT – Generative Pre-trained Transformer – генеративний попередньо натренований трансформер;

LSTM – Long Short-Term Memory – довготривала короткочасна пам'ять;

NLP – Natural Language Processing – обробка природної мови;

NSP – Next Sentence Prediction – прогнозування наступного речення;

RNN – Recurrent Neural Network – рекурентна нейронна мережа.

ВСТУП

У сучасному світі соціальні мережі відіграють ключову роль у комунікаціях між людьми, надаючи платформи для обміну інформацією, думками та досвідом. Соціальні мережі – це інтерактивні технології, що сприяють обміну, збереженню та створенню інформації [1]. З розвитком технологій обробки природної мови та штучного інтелекту з'являються нові можливості для оптимізації та покращення цих комунікацій. Однією з таких можливостей є використання перефразування повідомлень у соціальних мережах.

Повідомлення – це передача інформації між двома або більше людей. Зміст повідомлення може бути різноманітним, вичерпним [2]. Основною характеристикою повідомлень є канал передачі, який і відрізняє повідомлення від інших методів передачі інформації. Канал передачі може бути як і поштовим листом, так SMS чи електронним листом. Задля поліпшення процесу обміну інформацією потрібно використовувати перефразування.

Перефразування тексту може використовуватися для багатьох цілей, включаючи підвищення ясності повідомлень, уникнення плагіату [3], адаптацію змісту до різних аудиторій або навіть для автоматизованого створення унікального контенту у соціальних мережах. В контексті соціальних мереж, де швидкість та точність передачі інформації мають велике значення, здатність ефективно перефразувати короткі повідомлення може значно покращити взаєморозуміння та взаємодію між користувачами.

Актуальність перефразування повідомлень обумовлена зростаючою потребою в інструментах для поліпшення комунікацій в соціальних мережах. Це може стати ведучою силою у покращенні спілкування між людьми та створенням якісного контенту. Розробка таких інструментів

може сприяти на зростання різноманітного та доступного контенту в інтернеті.

Проблематика цього дослідження є у тому, що розроблені інструменти та алгоритми мають зберігати оригінальний зміст, сенс та емоційне забарвлення, повинні розуміти контекст, генерувати текст, який є читабельним, зрозумілим, цікавим та не є плагіатом інших авторів. Вирішення таких проблем вимагають розуміння лінгвістичних правил, знання мови та розуміння алгоритмів обробки природної мови.

Мета даної роботи є аналіз існуючих методів перефразування тексту, оцінка ефективності їхньої роботи та можливі ускладнення під час імплементації у популярні соціальні мережі. Буде оцінено найпопулярніші методи перефразування та моделі, а також оцінено роботу за допомогою мови програмування Python. Таким чином, буде проаналізовано декілька моделей та методів.

Підсумовуючи написане, дане дослідження має значення для подальшого розвитку обміну технологій та соціальних мереж. За допомогою використання методів штучного інтелекту можна покращити якість, швидкість та ефективність обміну інформацією між людьми.

Результати цього дослідження мають потенціал не лише для наукового використання, але й для практичного застосування в розробці соціальних мереж, відкриваючи нові перспективи для розробників та користувачів соціальних платформ. Тому зусилля, спрямовані на вивчення та вдосконалення методів перефразування, є вкрай важливими для досягнення прогресу в цій області.

1 АНАЛІЗ ПРЕДМЕТНОЇ ГАЛУЗІ

1.1 Актуальність соціальних мереж

Для сучасної людини існування соціальних мереж не є новинкою: перша соціальна мережа – Classmates.com – з'явилася у 1995 році [4], а найвідоміша у світі мережа – Facebook – з'явилася у 2004 році [5]. На даний момент користувачами додатків зі спілкуванням налічуються 5.04 мільярди осіб, або 63.3% користувачів. В Україні інтернет користувачів налічується 29.47 мільйонів, з них 25 мільйонів зареєстровані хоча б в одній з соціальних мереж [6].

Соціальна мережа – це онлайн-платформа або додаток завдяки якому люди щодня можуть обмінюватися інформацією [7]. Завдяки ним люди можуть створювати сторінки з власною інформацією, знаходити друзів та створювати мережі, писати один одному повідомлення, ставити лайки або дизлайки, репостити записи, створювати, редагувати чи видаляти записи, ділитися відео, фото або аудіо форматами. Функціонал соціальних мереж необмежений. Користувачі можуть знаходитися у всесвіті, де їм все відомо: який актор виграв нагороду або в кого зі знайомих нова покупка.

Соціальні мережі можна поділити на наступні категорії:

- месенджери;
- сайти або додатки для знайомств;
- блоги;
- обмін фотографіями та відео;
- професійні та тематичні;
- форуми.

За даними Shopify найпопулярнішими за кількістю активних користувачів є наступні соцмережі:

- Facebook – 3,065 мільйонів;
- Youtube – 2,504 мільйонів;

- Instagram – 2,000 мільйонів;
- WhatsApp – 2,000 мільйонів;
- TikTok – 1,582 мільйонів;
- WeChat – 1,343 мільйонів;
- Facebook Messenger – 1,010 мільйонів;
- Telegram – 900 мільйонів.

Статистику використання додатків наведено на рисунку 1.1.

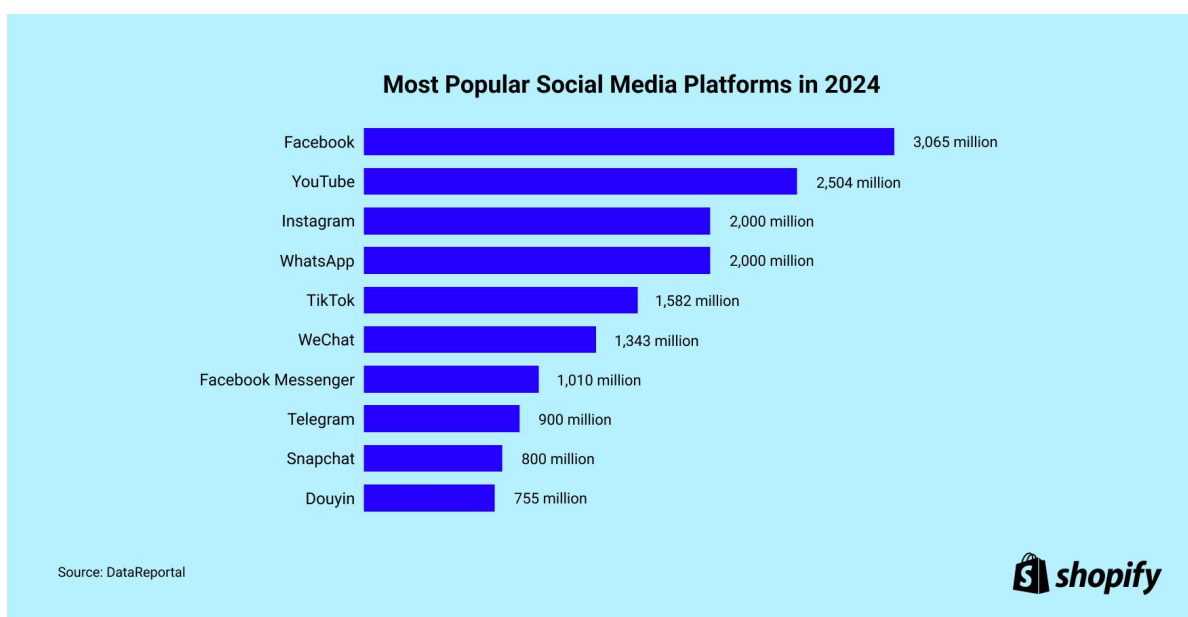


Рисунок 1.1 – Статистика найпопулярніших соціальних мереж станом на 2024 рік за Shopify [8]

Соціальні мережі можуть збирати разом велику кількість людей: можна знайти групу за інтересами та знайти там однодумців. Можна збирати людей на перегляд відео трансляцій. Таким чином, деякі люди або бренди можуть впливати на велику кількість людей - підписників, тому їх називають інфлюенсерами.

Інфлюенсери – це користувачі соціальних мереж, у який є вплив на власну аудиторію. Неважливо, яка кількість підписників в людини або компанії або ким вони є: якщо є група людей, які будуть довіряти думці

користувача, це на руку йому і рекламодавцям. Мета створення цього впливу є продаж реклами за засобом створення власного бренду, якому будуть довіряти люди.

Кожен вислів таких людей має значний вплив на інформаційне поле, але не завжди є зрозумілим, який посил несе висловлювання людей у соціальних мережах, тому перефразування тексту для покращення зрозумілості може бути корисним.

Перефразування також може використовуватися задля надання повідомленню більш доброзичливого змісту або привабливості для власної аудиторії. Таким чином, особистий бренд може зростати за допомогою створення вигляду, що людина розуміється на темі або симпатизує власним підписникам. Перефразування може впливати на залучність аудиторії у випадках, коли читачам відгукується те, що було написано. Потрібно у таких випадках знати контекст, знати цільову аудиторію інфлюенсера. Вдалих жарт на сторінці, на якій основна частина підписників полюбляє чорний гумор викличе більше взаємодії і більше активності, аніж цей самий жарт на сторінці, де зібрані люди, у який цей жарт матиме образливий відтінок. Вдалі повідомлення та контент на сторінці, які допоможуть сформувати успішний образ, — це вірний шлях на зростанні власного бренду.

Але важливо зазначити, що не сама ідея може стати запорукою зросту бренду, як її подача. Саме подача є запорукою поширення контенту у соціальних мережах. До подачі відноситься якість інформації. Повідомлення повинно мати вірусну ідею, якісну подачу та правильно обраний час для публікації.

Таким чином, саме перефразування представлення повідомлення може стати ключем для вірусності посту. Красиво сформоване речення є запорукою для поширення контенту між користувачами соціальних мереж.

Брендові облікові записи також мають на увазі, що публікація поста в одній країні може бути успішним, в іншій — недоречним або образливим.

Важливо підлаштовуватися під інші країни, в якій бренд хоче просуватися, також під різні мови. Перефразування є особливо важливим для світових брендів, які від настрою аудиторії можуть втрачати довіру до компанії. Воно дозволяє не просто перекласти речення, але й адаптувати під мову та країну, для якої має бути ця публікація, передаючи сенс та почуття, вкладенні в повідомлення, та не використовуючи образливі або неприємні для нації образи.

Вплив мають і повідомлення, які пишуть один одному користувачі таких платформ. Часто в переписках використовуються скорочення, сленг або настрій, який співрозмовник не може розпізнати, тому перефразування повідомлень є гарним інструментом задля покращення розуміння в комунікації.

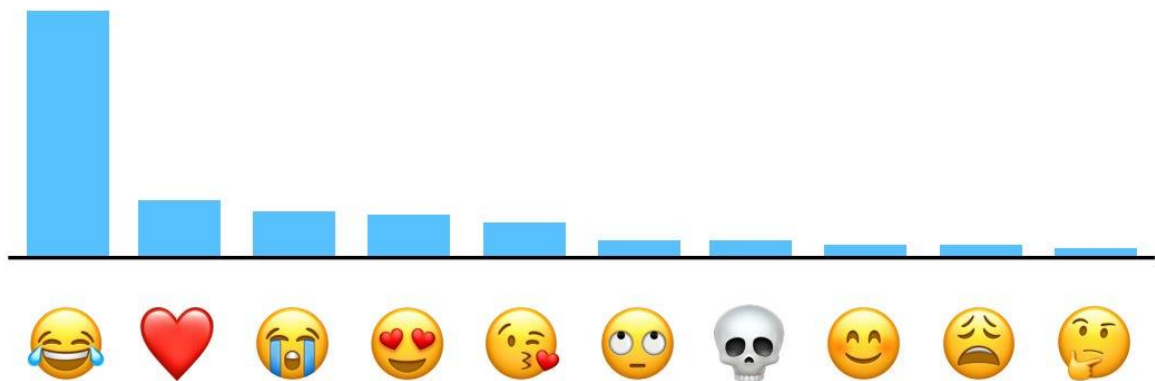
Не завжди зрозуміло, що має співрозмовник на увазі, пишучи коротке повідомлення, або що означає скорочення, яке він тільки що використав.

Особливо важким для розуміння такі деталі є літнім людям, які користуються соціальними мережами для спілкування з родиною. Такі повідомлення важко перефразувати та пояснювати у режимі переписування.

Перефразування може допомогти зрозуміти контекст, емоцію, посил який хотів передати автор.

Також, це може бути корисним для ділових співбесід, брейнштормів, на яких люди пишуть короткі повідомлення через нестаток часу або для спілкування між друзями, коли не зовсім зрозуміло використовуються терміни, емодзі тощо.

Емодзі, стікери, GIFки – це невеликі анімовані або статичні малюнки чи відео, які несуть якісь емоції чи посил в собі. Використання таких засобів спілкування вже давно є популярним, бо це не займає велику кількість часу та робить чат або пост емоційніше. Статистику використання емодзі можна побачити на рисунку 1.2.



The Count Mean Sketch technique allows Apple to determine the most popular emoji to help design better ways to find and use our favorite emoji. The top emoji for US English speakers contained some surprising favorites.

Рисунок 1.2 – Найпопулярніші у використанні емодзі [9]

З цього походять наступні елементи повідомлення – мему. Мему – це культурні елементи, до яких відносяться зображення, відео, аудио, тексти та фрази, які стають популярними та вірусними в інтернет-просторах. Термін "мем" був введений британським біологом Річардом Докінзом у книзі "Егоїстичний ген" для опису ідеї або поведінки, яка поширюється від однієї людини до іншої в культурі [10]. Зараз люди вже не жартують анекдоти, вони діляться мему, що ідеалістично є більш вдалим способом обміну інформацією між користувачами. Наприклад, будь яка людина може створити мем, спеціально або випадково, і методів для його створення набагато більше, ніж в анекдотів. Не на кожен життєву ситуацію знайдеться вдалий анекдот, як знайдеться мем. Мему несуть в собі культурний образ, вони мають підтекст, який кожен може зрозуміти по своєму. Розвиток цієї культури став потужним товчком для поширення інформації.

Приклад мему у вигляді зображення наведено на рисунку 1.3.



Рисунок 1.3 – Приклад мему з постіронією

Перефразування мемів є запорукою розуміння між різними культурами, в яких відсутнє розуміння цінностей між одним та іншим.

Таким чином, перефразування повідомлень може стати інструментом для адаптації в різних культурних умовах та цінностях, може забезпечити вірусність у соціальних мережах, стати в нагоді у спілкуванні людей різних вікових категорій та допомогти забезпечити репутацію бренду чи інфлюенсеру. Необхідно виправляти зміст, посил, втілення ідеї, помилки, перевіряти текст на культурну відповідність та сумісність з контекстом.

1.2 Огляд існуючих методів перефразування

Перефразування текстів є важливим аспектом в багатьох сферах, від перефразування повідомлень у приватних повідомленнях до публічних публікацій, де вміння показати та втілити ідею різними способами може значно покращити її сприйняття аудиторією та поширення між користувачами соціальних мереж. Існують різні методи та техніки перефразування, кожен з яких має свої переваги та обмеження.

Ручне перефразування – це процес, коли проста людина або експерт змінює текст, щоб висловити ту саму ідею іншими словами. Цей процес

вимагає від людини розуміння контексту повідомлення, глибокого розуміння теми та контексту і знання мовних конструкцій. Особливістю цього методу є більш чітке розуміння живої мови і здатність передати людські емоції. Такий процес займає велику кількість часу, але натомість надає гарний результат. Але такий метод потребує людину-експерта, яка буде перекладати або адаптувати текст.

Зворотній переклад – це процес, який може бути і автоматизованим, і людським. Задача такого методу перекласти текст на одну мову і потім назад на оригінальну, і під час перекладу можна використовувати багато мов. Зараз у людей в наявності багато інструментів для перекладання текстів, і наявність фахівця не обов'язкова, навпаки, завдяки такому перекладу може втратитися контекст або текст може адаптуватися під інші мови.

Наприклад, переклад речення: “Навіть у найтемнішу ніч зорі світять яскравіше, нагадуючи нам про надію, яка завжди є, навіть коли її не видно.”

Через одну мову: “Навіть найтемнішої ночі зірки світять яскравіше і нагадують нам про надію, яка завжди є, навіть коли її не видно.”

Через три мови: “Навіть у найтемнішу ніч зірки сяють яскравіше, нагадуючи нам про надію, яка завжди всередині нас, навіть коли ми її не бачимо.”

Через 5 мов: “Навіть у найтемніші ночі зірки яскраво сяють, нагадуючи нам про надію, яка завжди всередині нас, навіть коли ми її не бачимо.”

Можна побачити, як змінюється текст особливо не змінюючи сенсу першого речення і не втрачаючи власної актуальності. Цей метод займає багато часу і важкий для автоматизації, але може використовуватися для невеликих цілей. Він не враховує контекст і може змінити настрій, важко перекладає ідіоми, жарти та сленг, але виконує основну мету.

Використання існуючих систем синонимайзерів. Вони працюють на основі зміни слів на схожі за змістом, змінюючи речення, але не втрачаючи значення. Таким чином, можна зберігати сенс не втрачаючи основної думки. Беручи речення “Навіть у найтемнішу ніч зорі світять яскравіше, нагадуючи нам про надію, яка завжди є, навіть коли її не видно”, синонимайзер переклав її як “Навіть у найтемнішу годину зорі світять яскраво, нагадуючи про надію, що завжди присутня, навіть коли вона недоступна для очей”. Такі системи можуть стати в нагоді, коли необхідно скористуватися ними декілька разів, але вони займають багато часу для роботи.

Методи машинного навчання в штучному інтелекті представляють собою сучасний та ефективний підхід до перефразування. Використовуючи методи та алгоритми глибинного навчання можна генерувати нові повідомлення на основі контексту та історії старих повідомлень. Вони враховують більш точний зміст та контекст, зберігають мотиви у тексті, можуть додати чи замінити сенс тощо. На даний момент популярність трансформерів стає більшою за можливість використання різноманітних функцій для генерації коду. Маючи невеликі зразки історії повідомлень автора можна зберегти унікальний авторський стиль з використанням сленгу, помилок чи мовних оборотів, яким користується чи користувалася дана людина. Такі моделі самонавчаються на датасетах, що може гарно впливати на якість роботи в подальшому.

Але не дивлячись на увесь функціонал та можливості, які надають дані моделі, потрібно і звертати увагу на окремі фактори, які можуть призвести до зниження якості роботи. Відсутність мовних даних для навчання, можливості визначати контекст для перефразування, нечітко визначена задача та старі датасети для навчання можуть призвести до збільшення помилок у генерації та перефразуванні повідомлень. Вони не можуть прораховувати незвичайні контексти, використовувати нові сленгові слова та розуміти, який стиль буде найбільш відповідним для

повідомлень. Після такої генерації зазвичай необхідна перевірка для визначення та виправлення помилок у тексті.

1.3 Технології машинного навчання у перефразуванні

Для використання систем штучного інтелекту у питаннях перефразування текстів потрібно звернутися до NLP або Natural Language Processing. Обробка натуральних мов - це галузь штучного інтелекту та в особливості лінгвістики, яка займається взаємодією між людською мовою та її обробкою машинним кодом [11]. Зазвичай до цієї галузі відносять інтерпретація та генерація мови комп'ютерами. В обробці натуральних мов виконуються токенізація мови, обробка тексту, аналіз побудови речення, граматичних основ, контексту, знаходження об'єктів та суб'єктів, переклад на інші мови та емоційний аналіз. Зазвичай NLP використовуються для створення чат-ботів, розумних перекладачів, аналізаторів спаму, пошукових систем та багато іншого, де необхідна обробка людської мови.

Використання NLP у розробці програмного забезпечення зараз дуже поширене. Будь-яка хоча бути оптимізованою та прискореною.

У додатках використання NLP допомагає при перевірці тексту, генерації, перекладу. Таким чином, кожен додаток стає кращим, більш автоматизованим. Використання технологій штучного інтелекту може значно покращити програмне забезпечення, адже впливає на якість, швидкість та оптимізацію обміну інформацією.

Навіть у соціальних мережах вже є перекладачі зі штучним інтелектом або генерація тексту. Дані технології відкривають нові можливості для розробки і розвитку програмного забезпечення, роблячи його більш ефективним та клієнт орієнтованим.

Приклади використання обробки натуральної мови наведено на рисунку 1.4.

Applications of Natural Language Processing

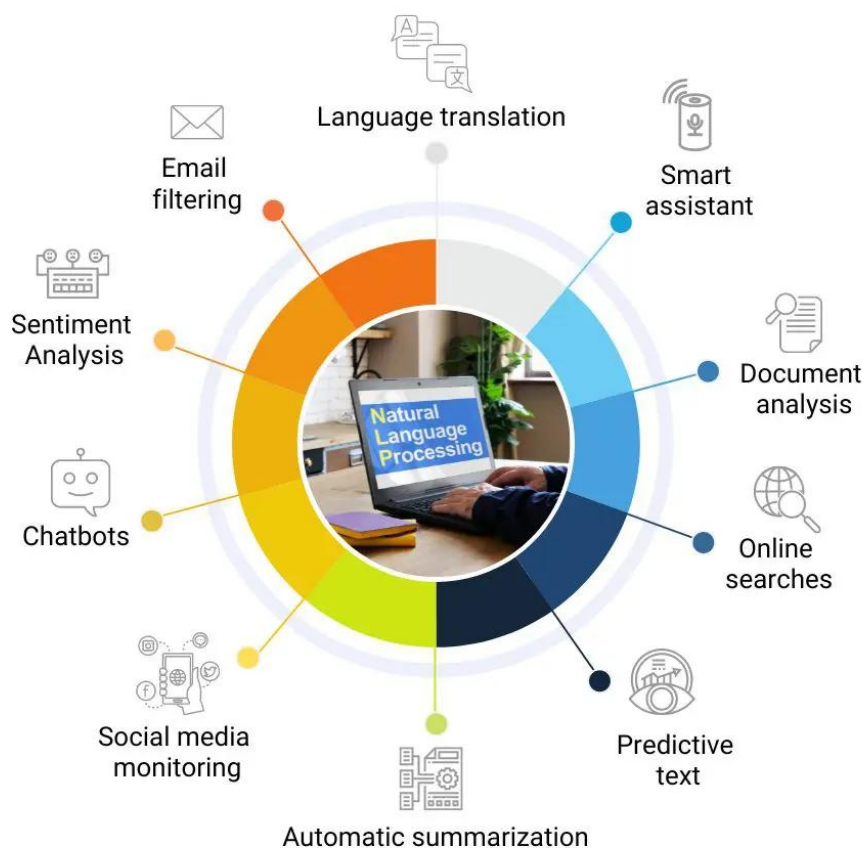


Рисунок 1.4 – Використання обробки природної мови [12]

У сучасних технологіях перефразування тексту велике місце займають методи машинного навчання, особливо моделі, засновані на архітектурі трансформерів. Ці моделі відіграють вирішальну роль у розвитку автоматизованого перефразування, пропонуючи передові рішення, здатні враховувати широкий контекст та зберігати нюанси первісного повідомлення.

Трансформери – це тип архітектури нейронних мереж, який був представлений у 2017 році компанією Google [13]. Трансформери значно покращили можливості NLP генерувати та обробляти текст.

У трансформерів є власні характеристики, які відрізняють їх від інших моделей. В них є механізм уваги, який дозволяє визначати коефіцієнт важливості слів у контексті генерації тексту. Це дозволяє контролювати генерацію. Таким чином у довгому тексті не втрачається первинний сенс генерації. Також трансформери можуть обробляти одразу усі слова в тексті, що відрізняє їх від RNN - рекурентних нейронних мереж. Це впливає на швидкість навчання та роботи нейронної мережі.

Архітектура Енкодер-Декодер дозволяє обробляти трансформеру текст швидко, адже енкодер транслює вхідний текст як внутрішнє представлення для мереж, а декодер з цього робить вихідний текст. Архітектура трансформерів масштабована та адаптована до великого обсягу моделей для навчання. Приклад схеми трансформера наведено на рисунку 1.5.

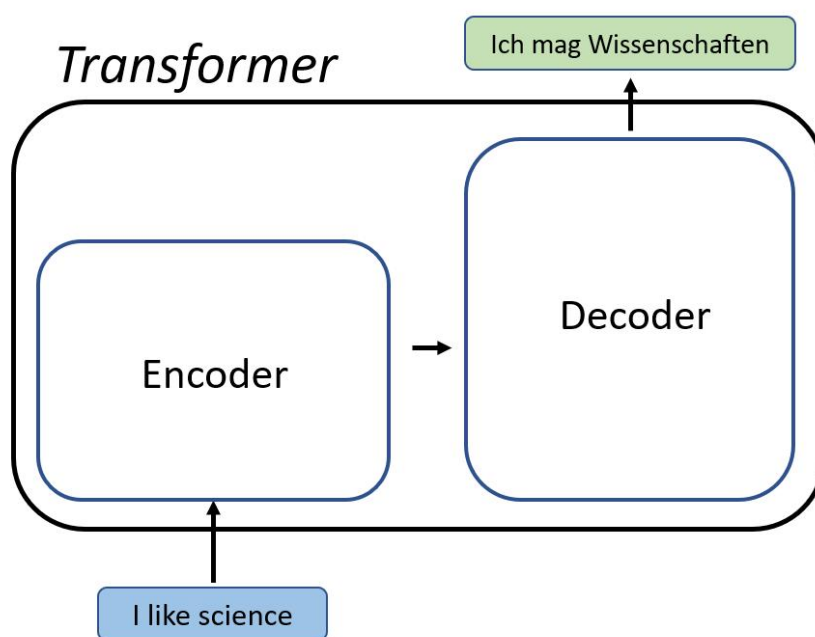


Рисунок 1.5 — Схема роботи трансформера

Відомі моделі трансформерів були представлені провідними компаніями в штучному інтелекті, такими як Google та OpenAI.

Найвідомішими з них на даний момент стали BERT від Google, GPT від OpenAI та T5 або Text-to-Text Transfer Transformer від Google.

При розробці трансформерів науковці відмовилися від стандартного підходу який раніше використовувався у моделях RNN та CNN, на користь використання механізму уваги, що значно покращує швидкість генерації тексту. Такі моделі можна використовувати для перефразування за рахунок можливості обробки великої кількості тексту. Вони можуть перефразувати текст у різних стилях написання та оформлення, використовуючи новий підхід.

Але навчання таких великих моделей займає багато ресурсів, які важко забезпечити навіть великим компаніям. Питання безпеки роботи таких моделей також є актуальним. Забезпечення якісними даними для навчання є складною задачею для розробників та науковців.

1.4 Виклики та обмеження методів перефразування

Названі методи перефразування, незалежно від того, чи вони засновані на ручних зусиллях або технологіях машинного навчання, стикаються з рядом викликів та обмежень. Складнощі у використанні даних методів викликані багатьма недоліками роботи перерахованих моделей та методів.

Поставлена задача є запорукою вибору метода для перефразування. Коли необхідно змінити стиль написання, коли потрібно виправити орфографічні помилки, а коли треба змінити емоційний фон повідомлення – від цього залежить обраний метод. Також як часто і для яких цілей буде використовуватися модель чи метод перефразування.

Одним з важливих викликів є збереження авторського права. У моментах коли перефразування використовується задля плагіату контенту потрібно знаходити вирішення проблеми. Адже перефразування може зіштовхнути користувача з наступними моментами, а саме юридичними.

Крадіжка авторського права або порушення прав користування платформою через перефразований контекст. Саме тому краще перевіряти згенерований контент.

Збереження значення та змісту оригінального тексту також являється викликом для перефразування. При зміні слів, мовних конструкцій можлива втрата ідеї та змісту початкового тексту. Особливо, якщо текст мав у собі складні та незрозумілі для всіх ідеї. Наприклад, якщо взяти назву пісні співачки Agnes Obel “Run Cried the Crawling”, яку за словами виконавиці не усі носії мови можуть перекласти. Вона інтерпретує назву як “”Біжи”, – кричав Повзучий”. При використанні існуючих сервісів можна отримати наступні перефразовані варіанти: “Cry Out, the Crawling”, “A run called a crawl” або “Run and crawl together”. В усіх варіантах перефразування втрачається оригінальний сенс. В цьому випадку складно перефразувати незвичну мовну конструкцію. “Don't count your chickens before they hatch” - одне з популярних прислів'їв в англійській мові. Prior to hatching, avoid counting your hens — це перефразоване прислів'я. Воно зберігає сенс при використанні перефразування, але аудиторія може просто сприйняти його як звичайне речення, а не словесний оборот. З цього можна зробити висновок — текст не існує у вакуумі. Окрім тексту його сенс визначають емоційне забарвлення, контекст, автор, культура та аудиторія. Автоматизовані перефразування не можуть враховувати такі аспекти, тому вони можуть зіткнутися з такими проблемами. Таким чином, перефразований текст просто втратить оригінальний сенс.

Текст існує як складова мови, а мова і культура – нерозривні поняття. Саме мова визначає культуру, а культура може інтерпретувати мову. Саме тому у світі існує як мінімум 5 версій англійської мови, не рахуючи діалекти і сленг, притаманний окремим групам. На цьому моменті перефразування також може стикнутися з купою проблем. Якщо не врахувати ці нюанси, можна втратити важливі культурні аспекти, ідіоми і отримати неправильну інтерпретацію тексту.

Усі ці нюанси стають перешкодами для розробки якісних програм для перефразування, що робить розробку коштовною і не доступною для багатьох компаній. Таким чином, індивідуальні розробники чи невеликі компанії просто можуть не вправитися з задачею.

1.5 Огляд існуючих перефразувачів

На даний момент існує багато сервісів-перефразувачів, якими можна скористатися онлайн.

Quill Bot: сервіс, який на даний момент підтримує перефразування на 26 мов, до числа яких входить українська та англійська. Можна обрати стиль написання, наприклад більш формальний або скорочений. Можна використовувати API для додавання сервісу у свій проект. Можна корегувати точність або креативність створеного тексту. Також у додатку йдуть перевірка на помилки, на генерацію штучного інтелекту, сумарайзер та підрахування відсотку цитування.

REPHRASE: сервіс, який дозволяє перефразувати 22 мовами світу та також має вибір стилю написання. Також має API, що дозволяє інтегрувати систем у нові проекти, плагіат перевірку та граматичну перевірку.

PARAPHRASER: сайт, що також підтримує 22 мови, але на відміну від інших, має платну розширену версію та безкоштовну обмежену. Але також він наводить статистику для тексту: яка кількість слів, який рівень мови для читання, час тощо. Також має API та додаткові версії на перевірку плагіату тощо.

Оглядаючи ще багато сайтів, можна побачити, що вони схожі один на одного. Кожен має однаковий набір функцій, додаткову розширену платну версію тощо.

1.6 Постановка задачі

Постановка задачі перефразування повідомлень або текстів у соціальних мережах може бути визначена як складний процес, що охоплює різні аспекти та цілі.

Цілі перефразування:

- Підвищення зрозумілості;
- виправлення граматичних та орфографічних помилок;
- адаптація до аудиторії;
- уникнення плагіату;
- застосування алгоритмів глибокого навчання.

Підвищення зрозумілості необхідно для спрощення тексту, щоб він став більш зрозумілим для ширшої аудиторії. Виправлення помилок також підвищує читаємість та робить аудиторію більш лояльною. Зміна стилістики тексту для адаптації до певної цільової аудиторії, враховуючи її вік, культурні особливості тощо. Перефразування з метою створення нового унікального контенту на основі вже існуючого матеріалу, зберігаючи при цьому його первісний зміст.

Потрібно проаналізувати існуючі моделі для перефразування текстів, щоби оцінити можливості роботи та з якими проблемами можна зіштовхнутися при роботі з ними.

При визначенні вимог до дослідження методів перефразування коротких повідомлень для соціальних мереж необхідно враховувати ряд ключових аспектів, щоб забезпечити якість і релевантність отриманих результатів. У дослідженні розглядаються наступні етапи:

- Аналіз роботи існуючих алгоритмів глибокого навчання;
- підвищення зрозумілості тексту у соціальних мережах;
- зміна стилю написання тексту;
- виправлення помилок.

2 ДОСЛІДЖЕННЯ МЕТОДІВ ПЕРЕФРАЗУВАННЯ

Для дослідження методів перефразування повідомлень у соціальних мережах будуть використовуватися існуючі методи глибинного навчання та оцінюватися наступні критерії: збереження оригінального сенсу, виправлення помилок, зміна стильового написання. Це основні критерії для перефразовування.

Повідомлення на якому будуть перевірятися моделі: “Hey their, Hope you're doing well! I just wanted to shear some exciting news about our upcoming project. We've made a lot of progress over the past few months, and I thought you'd like to here about it.

First off, our team has wrapped up the initial faze of the project, which involved a lot of research and data colletion. We've gained some really valuable insights that will help guild our next steps. Now, we're moving into the developepment phase where we'll start implementing all the strategies we've come up with.

Also, I'd love for you to join a meeting next Monday at 10 AM to chat about the project's progress and get your feedback. Your input is super inportant to us, and we think it will be instrumental in making this project a success.

Lastly, I just want to say thanks for your continued support and dedication. Together, we've got the potential to achieve some amazing results, and I'm confident that our combined efforts will lead to a succsesful outcome.

Looking forward to seeing you at the meeting!”.

У реченні спеціально допущено велику кількість помилок задля перевірки на перефразування, адже одна із мет - це перевірити на можливість виправлення помилок. Таким чином, використання перефразування має покрити базові необхідності користувача для того, щоб покращити читабельність тексту, підвищити лояльність або надати перефразованому тексту інший стиль.

Також буде перевірятися вміння зберігати сенс складних мовних конструкцій. Усі експерименти будуть проводитися на англійськомовних повідомленнях тому що на даний момент усі існуючі моделі найкраще працюють з англійською мовою.

Також для перевірки збереження складних мовних конструкцій буде використано приклад наступного повідомлення: "Despite the ostensibly insurmountable challenges posed by the intricate and multifaceted nature of the project, the team, drawing upon their extensive collective expertise and unwavering determination, managed to devise a comprehensive strategy that not only addressed the immediate issues but also anticipated potential future obstacles, thereby ensuring a sustainable and resilient solution".

При роботі будуть аналізуватися якість, правильність і точність роботи моделей та методів, які будуть використовуватися. Відповіді, що вони будуть давати, мають бути читабельними і мати властивості, які набуваються після перефразування. При кожному підході будуть враховуватися також легкість імплементації та швидкість.

2.1 Трансформери

Першими будуть перевірятися моделі-трансформери. Трансформери – це архітектура глибокого навчання, яка використовує механізм уваги. Вперше була представлена компанією Google Brain у 2017 році у журналі "Attention is all you need" [14]. Вони змінили уявлення про обробку природної мови. У даний момент вони гарно замінюють рекурентні нейронні мережі та довгі короткострокові пам'яті моделі(LSTM). Вони беруть за основу ідею механізм уваги, який однаково оцінює увесь текст не дивлячись на відстань між ним. Тому це спрощує залежність слів у тексті один від одного.

Механізм уваги – це компонент трансформерів, який дозволяє фокусуватися на усіх частинах вхідних даних для покращення роботи.

Основа ідея механізму у тому, щоб зважувати різні частини вхідного сигналу вибірково. Тому при генерації модель може звертати увагу на усі частини входу. Такий підхід полягає у зважуванні кожного токена чи слова за допомогою вагів. Приклад роботи механізму уваги наведено на рисунку 2.1.

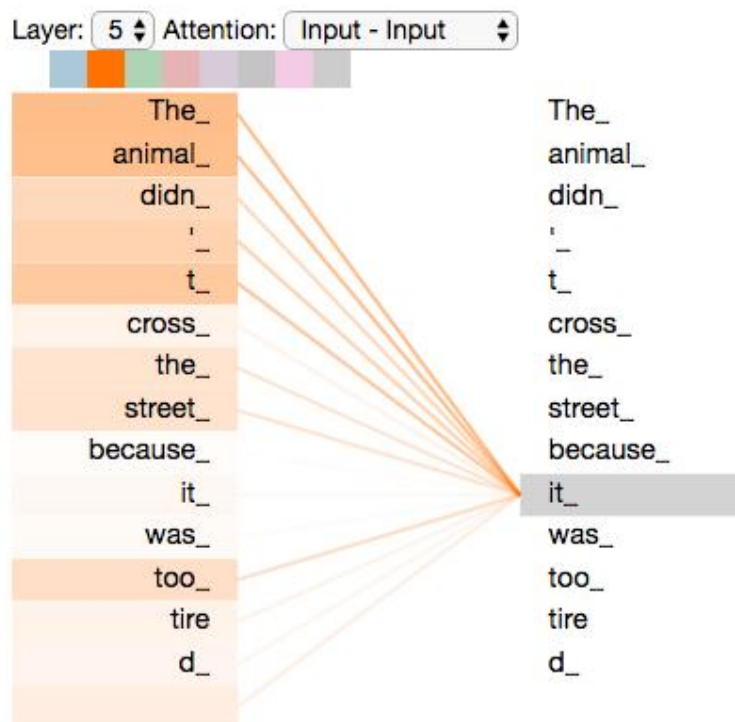


Рисунок 2.1 – Приклад роботи механізму уваги

Також є декілька видів уваги: самоскерована та мультикерована увага. У самоскерованій усі токени розраховуються вагою, незалежно від їх позицій. У мультикерованому підході розраховують декілька головних елементів уваг, де вони потім сплюсуються.

Таким чином трансформери за допомогою механізму уваги можуть обробляти складні залежності між токенами, знаходячи зв'язок і не втрачаючи контексту. У трансформерах використання механізму уваги використовується і в енкодерах, і в декодерах. Енкодери використовують самоувагу для вхідних даних, щоб зловити увесь контекст повідомлення, а

декодери використовують увагу для генерації вихідних даних. Деякі моделі-трансформери використовують механізм уваги у двонаправленому напрямку, враховуючи токени зліва і справа водночас. Таким чином, для токена розраховується увесь контекст. Це дозволяє збільшити кількість виконуваних задач та прискорити навчання і роботу моделі.

Деякі моделі-трансформери само навчаються, тому модель постійно розвивається, хоча інколи вони знаходять паттерни поведінки людей і стають менш ефективними. Трансформери – універсальні, тому підходять для великої кількості задач. Тому такі моделі-трансформери стали популярними і швидко знайшли свою нішу у продуктах штучного інтелекту.

На даний момент існують наступні моделі:

- BERT (Bidirectional Encoder Representations from Transformers) або нова назва GEMINI;
- GPT (Generative Pre-trained Transformers);
- T5 (Text To Text Transfer Transformers);
- RoBERTa (Robustly Optimized BERT Pretraining Approach);
- XLNet;
- DistilBERT;
- ALBERT (A Lite BERT);
- ELECTRA;
- Switch Transformer.

2.1.1 BERT

BERT (GEMINI) модель була запропонована в книзі "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding" від авторів Джейкоба Девлінга, Мінг-Вей Чанга, Кентона Лі та Кристина Тутановой. Це двонаправлений трансформатор, який попередньо навчений за допомогою комбінації маскованої цілі, моделюванні мови і навчений на

датовані, що створений на основі Toronto Corpus Book і Вікіпедії [15]. BERT розшифровується як Bidirectional Encoder Representations from Transformer. Ця модель створена для попереднього навчання глибоких двонаправлених виглядів на основі не маркованого тексту шляхом спільного користування лівого та правого контекстів слів на усіх рівнях. BERT є концептуально простим та емпірично потужним.

Для роботи з BERT моделлю необхідно в коді вписувати промпт. Промпт – це запит для моделі, що саме вона має робити.

У першому випадку є перевірка на виправлення помилок у тексті зі збереженням основного змісту та мовних конструкцій. При роботі видно, що модель впоралася з усіма помилками і зберегла конструкцію повідомлення. Результат роботи можна побачити у таблиці 2.1.

Таблиця 2.1 – Результат роботи виправлення помилок з BERT

Вхідне повідомлення	Вихідне повідомлення
<p>Hey their,</p> <p>Hope you're doing well! I just wanted to shear some exciting news about our upcoming project. We've made a lot of progress over the past few months, and I thought you'd like to here about it.</p> <p>First off, our team has wrapped up the initial faze of the project, which involved a lot of research and data colletion. We've gained some really valuable insights that will help guild our next steps. Now, we're moving into the developepment phase where we'll start implementing all the strategies</p>	<p>Hey there,</p> <p>Hope you're doing well! I just wanted to share some exciting news about our upcoming project. We've made a lot of progress over the past few months, and I thought you'd like to hear about it.</p> <p>First off, our team has wrapped up the initial phase of the project, which involved a lot of research and data collection. We've gained some really valuable insights that will help guide our next steps. Now, we're moving into the development phase where we'll start implementing all the strategies</p>

Продовження таблиці:

Вхідне повідомлення	Вихідне повідомлення
<p>we've come up with.</p> <p>Also, I'd love for you to join a meeting next Monday at 10 AM to chat about the project's progress and get your feedback. Your input is super important to us, and we think it will be instrumental in making this project a success.</p> <p>Lastly, I just want to say thanks for your continued support and dedication. Together, we've got the potential to achieve some amazing results, and I'm confident that our combined efforts will lead to a successful outcome.</p> <p>Looking forward to seeing you at the meeting!</p>	<p>we've come up with.</p> <p>Also, I'd love for you to join a meeting next Monday at 10 AM to chat about the project's progress and get your feedback. Your input is super important to us, and we think it will be instrumental in making this project a success.</p> <p>Lastly, I just want to say thanks for your continued support and dedication. Together, we've got the potential to achieve some amazing results, and I'm confident that our combined efforts will lead to a successful outcome.</p> <p>Looking forward to seeing you at the meeting!</p>

У другому варіанті перевіряється перефразування на професійний та діловий стиль тексту, при збереженні основного сенсу та зміні слів і мовних конструкцій. Видно, що модель впоралася, але необхідно заповнити пусті пропуски, що може стати проблемою при автоматизованому використанні моделі. Результат можна побачити у таблиці 2.2.

Таблиця 2.2 – Результат роботи перефразування у діловому стилі з BERT

Вхідне повідомлення	Вихідне повідомлення
<p>Hey there,</p> <p>Hope you're doing well! I just wanted to share some exciting news about our upcoming project. We've made a lot of progress over the past few months, and I thought you'd like to hear about it.</p> <p>First off, our team has wrapped up the initial phase of the project, which involved a lot of research and data collection. We've gained some really valuable insights that will help guide our next steps. Now, we're moving into the development phase where we'll start implementing all the strategies we've come up with.</p> <p>Also, I'd love for you to join a meeting next Monday at 10 AM to chat about the project's progress and get your feedback. Your input is super important to us, and we think it will be instrumental in making this project a success.</p> <p>Lastly, I just want to say thanks for your continued support and dedication. Together, we've got the potential to achieve some amazing</p>	<p>Subject: Update on [Project Name] and Invitation to Meeting</p> <p>Dear [Name],</p> <p>I hope this email finds you well. I'm writing to share some positive developments on our upcoming project, [Project Name].</p> <p>The team has successfully completed the initial phase, which focused on in-depth research and data collection. This has yielded valuable insights that will inform our next steps. We are now transitioning to the development phase, where we will begin implementing the strategies we formulated.</p> <p>To further discuss the project's progress and obtain your valuable feedback, I would be delighted if you could join a meeting next Monday at 10:00 AM. Your input is instrumental to the project's success.</p> <p>Thank you for your continued support and dedication. Together, we have the potential to achieve significant results, and I am confident that our combined efforts will lead to a</p>

Продовження таблиці:

Вхідне повідомлення	Вихідне повідомлення
<p>results, and I'm confident that our combined efforts will lead to a successful outcome.</p> <p>Looking forward to seeing you at the meeting!</p>	<p>successful outcome.</p> <p>Looking forward to seeing you at the meeting.</p> <p>Sincerely,</p> <p>[Your Name]</p>

Переклад з української на англійську зі збереженням основного сенсу та складних мовних конструкцій. Модель впоралася гарно і замінила складні мовні конструкції на відповідники в англійській мові. Результат продемонстрований у таблиці 2.3.

Таблиця 2.3 — Результат перекладу з BERT

Вхідне повідомлення	Вихідне повідомлення
<p>Зітхнувши важко, він усвідомив, що доля знову підклала йому свиню, змусивши пройти крізь терні до зірок, аби досягти омріяного успіху</p>	<p>With a heavy sigh, he realized that fate had once again thrown him a curveball, forcing him to navigate the thorny path to the stars in order to achieve his long-dreamed success.</p>

Він переклав деякі мовні конструкції, наприклад, “що доля знову підклала йому свиню” - "that fate had once again thrown him a curveball".

2.1.2 GPT

GPT – це Generative Pre-trained Transformer, основною версією є велика мовна модель GPT-3. Це модель-трансформер глибинної нейронної мережі, яка працює на механізмі уваги. Має 175 мільярдів параметрів,

кожен з яких має 16-бітну точність [16]. Продемонстрував гарну точність при роботі з багатьма мовами світу.

Використання GPT на даний момент:

- Використовується для GitHub Copilot, програмного забезпечення для генерації, завершення та виправлення програмного коду. Може використовуватися у різних редакторах коду та IDE;
- використовується у декількох продуктах Microsoft для перекладу людської мови у комп'ютерну;
- використовується у CodexDB для генерації запитів SQL коду;
- використовується у “Project December”;
- використана для написання 8 есе про нешкідливість ШІ для людства та ці есе були опубліковані в одну статтю The Guardian;
- використовувалася в грі AI Dungeon;
- припущено, що GPT може робити скрінінг хвороби Альцгеймера на ранніх стадіях.

Використання GPT у проектах обмежено політикою компанії OpenAI.

У 2024 році OpenAI випускає версію 4o, де “o” це omni, що означає “всеохоплюючий”. Така модель приймає на вхід будь-який формат, як текст, аудіо, зображення та відео і на виході генерує текст, аудіо та зображення. Ця модель може “спілкуватися”, швидко реагує на будь який тип входу.

Виправлення помилок у тексті для ChatGPT або API GPT з моделлю davinci має мати промпт — запит для роботи. Промпт буде наступним “Correct all mistakes in the following text”. Результат роботи можна побачити у таблиці 2.4.

Таблиця 2.4 — Результат виправлення помилок з GPT-3

Вхідне повідомлення	Вихідне повідомлення
Hey their, Hope you're doing well! I just	Hey there, Hope you're doing well! I just

Продовження таблиці:

Вхідне повідомлення	Вихідне повідомлення
<p>wanted to shear some exciting news about our upcoming project. We've made a lot of progress over the past few months, and I thought you'd like to here about it.</p> <p>First off, our team has wrapped up the initial faze of the project, which involved a lot of research and data colletion. We've gained some really valuable insights that will help guild our next steps. Now, we're moving into the developepment phase where we'll start implementing all the strategies we've come up with.</p> <p>Also, I'd love for you to join a meeting next Monday at 10 AM to chat about the project's progress and get your feedback. Your input is super inportant to us, and we think it will be instrumental in making this project a success.</p> <p>Lastly, I just want to say thanks for your continued support and dedication. Together, we've got the potential to achieve some amazing results, and I'm confident that our combined efforts will lead to a succsesful outcome.</p>	<p>wanted to share some exciting news about our upcoming project. We've made a lot of progress over the past few months, and I thought you'd like to hear about it.</p> <p>First off, our team has wrapped up the initial phase of the project, which involved a lot of research and data collection. We've gained some really valuable insights that will help guide our next steps. Now, we're moving into the development phase where we'll start implementing all the strategies we've come up with.</p> <p>Also, I'd love for you to join a meeting next Monday at 10 AM to chat about the project's progress and get your feedback. Your input is super important to us, and we think it will be instrumental in making this project a success.</p> <p>Lastly, I just want to say thanks for your continued support and dedication. Together, we've got the potential to achieve some amazing results, and I'm confident that our combined efforts will lead to a successful outcome.</p>

Продовження таблиці:

Вхідне повідомлення	Вихідне повідомлення
Looking forward to seeing you at the meeting!	Looking forward to seeing you at the meeting!

Усі помилки були виправлені без зміни тексту, що відповідає запиту для роботи.

Тепер потрібно замінити стиль написання на діловий за допомогою запиту “Change the writing style to business”. Результат роботи наведений у таблиці 2.5.

Таблиця 2.5 — Результат роботи перефразування у діловому стилі з GPT-3

Вхідне повідомлення	Вихідне повідомлення
<p>Hey there,</p> <p>Hope you're doing well! I just wanted to share some exciting news about our upcoming project. We've made a lot of progress over the past few months, and I thought you'd like to hear about it.</p> <p>First off, our team has wrapped up the initial phase of the project, which involved a lot of research and data collection. We've gained some really valuable insights that will help guide our next steps. Now, we're moving into the development phase where we'll start implementing all the strategies we've come up with.</p>	<p>Dear [Recipient's Name],</p> <p>I hope this message finds you well. I am writing to share some exciting updates regarding our upcoming project. Over the past few months, we have made significant progress, and I believe you will find the developments noteworthy. Firstly, our team has successfully completed the initial phase of the project, which involved extensive research and data collection. Firstly, our team has successfully completed the initial phase of the project, which involved extensive research and data collection. We have obtained valuable insights that</p>

Продовження таблиці:

Вхідне повідомлення	Вихідне повідомлення
<p>Also, I'd love for you to join a meeting next Monday at 10 AM to chat about the project's progress and get your feedback. Your input is super important to us, and we think it will be instrumental in making this project a success.</p> <p>Lastly, I just want to say thanks for your continued support and dedication. Together, we've got the potential to achieve some amazing results, and I'm confident that our combined efforts will lead to a successful outcome.</p> <p>Looking forward to seeing you at the meeting!</p>	<p>will effectively guide our next steps. We are now transitioning into the development phase, where we will begin implementing the strategies we have devised.</p> <p>I would like to invite you to a meeting next Monday at 10 AM to discuss the project's progress and to gather your feedback. Your input is extremely important to us, and we believe it will be instrumental in ensuring the success of this project.</p> <p>Lastly, I want to express my gratitude for your continued support and dedication. Together, we have the potential to achieve remarkable results, and I am confident that our combined efforts will lead to a successful outcome.</p> <p>I look forward to seeing you at the meeting.</p> <p>Best regards, [Your Name]</p>

Результатом є діловий стиль написання повідомлення, що відповідає на запит. Також є пропуски, які мають бути заповнені, але стиль професійний збережено акуратно.

Останнім запитом буде перекласти речення зі збереженням оригінального значення із запитом “Translate into English with preservation of the original content”. Результатом є таблиця 2.6.

Таблиця 2.6 - Результат перекладу на англійську з моделлю GPT-3

Вхідне повідомлення	Вихідне повідомлення
Зітхнувши важко, він усвідомив, що доля знову підклала йому свиню, змусивши пройти крізь терні до зірок, аби досягти омріяного успіху	Sighing heavily, he realized that fate had once again played a dirty trick on him, forcing him to go through trials and tribulations to achieve the success he had dreamed of.

Результат з перекладом також гарний. Він переклав без використання словесних конструкцій, але дослівно переведено гарно. При цьому дослівно переклав складні мовні конструкції.

2.1.3 T5

T5 (Text-To-Text Transfer Transformer) — це модель глибинної нейронної мережі, яка розроблена дослідниками з Google Research. Це трансформер, який призначено для роботи і вирішення великої кількості задач формату з тексту в текст. Це означає, що багато задач NLP, такі як класифікація тексту, переклади, заповнення пропусків, генерація тексту, виправлення помилок тощо, є у вигляді текстової задачі і відповідь буде текстовою [17].

T5 використовує одну і ту саму архітектуру для усіх задач. Вона доступна у декількох розмірах, щоб можна було обрати модель для роботи за власними ресурсами.

T5 може відповідати на питання, генерувати і виправляти текст, тому є гнучкою текстовою моделлю.

При проведенні експерименту і використанні моделі T5 було обрано модель найменшого розміру через нестачу ресурсів для роботи.

При виправленні помилок модель впоралася нормально, але зрізала більше половини тексту і були повторення слів. Результат у таблиці 2.7.

Таблиця 2.7 — Результат роботи виправлення помилок з T5

Вхідне повідомлення	Вихідне повідомлення
<p>Hey their,</p> <p>Hope you're doing well! I just wanted to shear some exciting news about our upcoming project. We've made a lot of progress over the past few months, and I thought you'd like to here about it.</p> <p>First off, our team has wrapped up the initial faze of the project, which involved a lot of research and data colletion. We've gained some really valuable insights that will help guild our next steps. Now, we're moving into the developepment phase where we'll start implementing all the strategies we've come up with.</p> <p>Also, I'd love for you to join a meeting next Monday at 10 AM to chat about the project's progress and get your feedback. Your input is super inportant to us, and we think it will be instrumental in making this project a success.</p>	<p>Hey there,</p> <p>Hope you're doing well!</p> <p>Hey there,</p> <p>Hope you're doing well! I just wanted to share some exciting news about our upcoming project.</p> <p>First off, our team has wrapped up the initial phase of the project, which involved a lot of research and data collection. I'd love for you to join a meeting next Monday at 10 AM to chat about the project's progress.</p>

Продовження таблиці:

Вхідне повідомлення	Вихідне повідомлення
<p>Lastly, I just want to say thanks for your continued support and dedication. Together, we've got the potential to achieve some amazing results, and I'm confident that our combined efforts will lead to a successful outcome.</p> <p>Looking forward to seeing you at the meeting!</p>	

При оформленні ділового стилю повідомлення можна побачити, що сталася така сама ситуація, що і з виправленням помилок – текст просто було зрізано і були повторення. Результат можна побачити у таблиці 2.8.

Таблиця 2.8 — Результат роботи перефразування на діловий стиль з T5

Вхідне повідомлення	Вихідне повідомлення
<p>Hey there,</p> <p>Hope you're doing well! I just wanted to share some exciting news about our upcoming project. We've made a lot of progress over the past few months, and I thought you'd like to hear about it.</p> <p>First off, our team has wrapped up the initial phase of the project, which involved a lot of research and data</p>	<p>Hey there,</p> <p>Hope you're doing well!</p> <p>Hey there,</p> <p>Hope you're doing well! I just wanted to share some exciting news about our upcoming project.</p> <p>First off, our team has wrapped up the initial phase of the project, which involved a lot of research and data collection.</p>

Продовження таблиці:

Вхідне повідомлення	Вихідне повідомлення
<p>collection. We've gained some really valuable insights that will help guide our next steps. Now, we're moving into the development phase where we'll start implementing all the strategies we've come up with.</p> <p>Also, I'd love for you to join a meeting next Monday at 10 AM to chat about the project's progress and get your feedback. Your input is super important to us, and we think it will be instrumental in making this project a success.</p> <p>Lastly, I just want to say thanks for your continued support and dedication. Together, we've got the potential to achieve some amazing results, and I'm confident that our combined efforts will lead to a successful outcome.</p> <p>Looking forward to seeing you at the meeting!</p>	

З перекладом модель зовсім не впоралася. Результат у таблиці 2.9.

Таблиця 2.9 — Результат роботи перекладу з української з T5

Вхідне повідомлення	Вихідне повідомлення
<p>Зітхнувши важко, він усвідомив, що доля знову підклала</p>	<p>усвдомив, о дол нову дклала ому свин, мусиви роти кр терн до</p>

Продовження таблиці:

Вхідне повідомлення	Вихідне повідомлення
Йому свиню, змусивши пройти крізь терні до зірок, аби досягти омріяного успіху	рок, аи дости омрноо усу.

T5 модель не показала себе як гарний інструмент для перефразування текстів та виправлення помилок. З деякими задачами вона впоралась, але дуже слабо.

2.1.4 RoBERTa

RoBERTa або Robustly Optimized BERT Approach – це модель глибинного навчання та (NLP), яка була розроблена дослідниками і розробниками Facebook AI. Вона є продовженням моделі BERT і покращує її продуктивність шляхом використання більшого обсягу даних [18].

Ця модель використовує кращі параметри та методики навчання, що дозволяє досягнути більшої продуктивності. Вона натренована на більшій кількості даних, ніж оригінальна модель. Головною зміною від попередньої моделі це відмова використання NSP або прогнозування наступного речення, адже дослідники виявили, що це не покращує генерацію.

Ця модель також працює з усіма задачами NLP. Займається класифікацією тексту, визначенням емоційності, перефразуванням тощо.

При виправленні помилок модель впоралася гарно, помилки відсутні. Результат у таблиці 2.10.

Таблиця 2.10 — Результат роботи виправлення помилок з RoBERTa

Вхідне повідомлення	Вихідне повідомлення
Hey their,	Hey there,

Продовження таблиці:

Вхідне повідомлення	Вихідне повідомлення
<p data-bbox="225 286 823 645">Hope you're doing well! I just wanted to share some exciting news about our upcoming project. We've made a lot of progress over the past few months, and I thought you'd like to hear about it.</p> <p data-bbox="225 674 823 1223">First off, our team has wrapped up the initial phase of the project, which involved a lot of research and data collection. We've gained some really valuable insights that will help guide our next steps. Now, we're moving into the development phase where we'll start implementing all the strategies we've come up with.</p> <p data-bbox="225 1252 823 1671">Also, I'd love for you to join a meeting next Monday at 10 AM to chat about the project's progress and get your feedback. Your input is super important to us, and we think it will be instrumental in making this project a success.</p> <p data-bbox="225 1700 823 2058">Lastly, I just want to say thanks for your continued support and dedication. Together, we've got the potential to achieve some amazing results, and I'm confident that our combined efforts will lead to a</p>	<p data-bbox="845 286 1444 645">Hope you're doing well! I just wanted to share some exciting news about our upcoming project. We've made a lot of progress over the past few months, and I thought you'd like to hear about it.</p> <p data-bbox="845 674 1444 1223">First off, our team has wrapped up the initial phase of the project, which involved a lot of research and data collection. We've gained some really valuable insights that will help guide our next steps. Now, we're moving into the development phase where we'll start implementing all the strategies we've come up with.</p> <p data-bbox="845 1252 1444 1671">Also, I'd love for you to join a meeting next Monday at 10 AM to chat about the project's progress and get your feedback. Your input is super important to us, and we think it will be instrumental in making this project a success.</p> <p data-bbox="845 1700 1444 2058">Lastly, I just want to say thanks for your continued support and dedication. Together, we've got the potential to achieve some amazing results, and I'm confident that our combined efforts will lead to a</p>

Продовження таблиці:

Вхідне повідомлення	Вихідне повідомлення
seeing you at the meeting!	successful outcome. Looking forward to seeing you at the meeting!

При перефразуванні необхідно використовувати маски - пропуски, які модель буде заповнювати. Це не найкращий метод перефразування, не виправляється стиль, а слова можуть змінитися на ті самі, що були до перефразування. В цілому модель впоралася. Під час використання можуть виникнути проблеми, адже якісне маскування може зробити тільки людина, а не алгоритм. Але це не найкращий варіант для використання. Результат у таблиці 2.11.

Таблиця 2.11 — Результат роботи перефразування на діловий стиль з RoBERTa

Вхідне повідомлення	Вихідне повідомлення
Hey their, Hope you're doing well! I just wanted to shear some exciting news about our upcoming project. We've made a lot of progress over the past few months, and I thought you'd like to here about it. First off, our team has wrapped up the initial <mask> of the project, which involved a lot of research and data collection. We've gained some really valuable insights that will help guide our next steps. Now, we're moving into the development phase where we'll	Hey their, Hope you're doing well! I just wanted to shear some exciting news about our upcoming project. We've made a lot of progress over the past few months, and I thought you'd like to here about it. First off, our team has wrapped up the initial phase of the project, which involved a lot of research and data collection. We've gained some really valuable insights that will help guide our next steps. Now, we're moving into the development phase where we'll

<p>start implementing all the strategies we've come up with. Also, I'd love for you to join a meeting next Monday at 10 AM to chat about the project's progress and get your feedback. Your input is super important to us, and we think it will be instrumental in making this project a success. Lastly, I just want to say thanks for your continued support and dedication. Together, we've got the potential to achieve some amazing results, and I'm confident that our combined efforts will lead to a successful outcome. Looking forward to seeing you at the meeting!</p>	<p>start implementing all the strategies we've come up with. Also, I'd love for you to join a meeting next Monday at 10 AM to chat about the project's progress and get your feedback. Your input is super important to us, and we think it will be instrumental in making this project a success. Lastly, I just want to say thanks for your continued support and dedication. Together, we've got the potential to achieve some amazing results, and I'm confident that our combined efforts will lead to a successful outcome. Looking forward to seeing you at the meeting!</p>
---	---

З перекладом модель також не впоралася. Тому модель можна використовувати, але є набагато кращі і сильніші моделі.

2.2 Бібліотека TextBlob

TextBlob – це бібліотека Python для обробки природної мови, яка надає розробнику простий інтерфейс для використання та імплементації у додатки та виконання базових задач NLP. Ця бібліотека спрощує роботу з текстовими даними, для яких є інструменти, що легко використовувати [19].

Ця бібліотека може аналізувати тональність тексту, визначаючи його як позитивний, негативний чи нейтральний. Може визначати граматичні категорії слів як частин речення, що може допомогти у специфічних завданнях. Вона може визначати назви як іменовані сутності: міста, імена,

організації тощо. Переклад з однієї мови на іншу. Зробити синтаксичний розбір, визначити граматичну структуру речення, виправляти орфографічні помилки, резюмувати текст.

Ця бібліотека легка в програмуванні, в роботі і надає багато корисних інструментів, які інколи необхідні у роботі з розробкою додатків.

Перекладу на українську мову вона не має, як і зміну стилю тексту, тому можна спробувати виправити помилки у тексті. Для цього на допомогу приходять метод `def correct()` з бібліотеки. Результат роботи у таблиці 2.12

Таблиця 2.12 — Результат роботи виправлення помилок з TextBlob

Вхідне повідомлення	Вихідне повідомлення
<p>Hey their,</p> <p>Hope you're doing well! I just wanted to shear some exciting news about our upcoming project. We've made a lot of progress over the past few months, and I thought you'd like to here about it.</p> <p>First off, our team has wrapped up</p>	<p>Key their, Hope you're doing well! I just wanted to hear some exciting news about our coming project. He've made a lot of progress over the past few months, and I thought you'd like to here about it. First off, our team has wrapped up the initial face of the project, which involved a lot of</p>

Продовження таблиці:

Вхідне повідомлення	Вихідне повідомлення
<p>the initial faze of the project, which involved a lot of research and data colletion. We've gained some really valuable insights that will help guild our next steps. Now, we're moving into the developepment phase where we'll start implementing all the</p>	<p>research and data collection. He've gained some really valuable insight that will help guild our next steps. Now, we're moving into the development phase where we'll start implementing all the strategics we've come up with. Also, I'd love for you to</p>

<p>strategies we've come up with.</p> <p>Also, I'd love for you to join a meeting next Monday at 10 AM to chat about the project's progress and get your feedback. Your input is super important to us, and we think it will be instrumental in making this project a success.</p> <p>Lastly, I just want to say thanks for your continued support and dedication. Together, we've got the potential to achieve some amazing results, and I'm confident that our combined efforts will lead to a successful outcome.</p> <p>Looking forward to seeing you at the meeting!</p>	<p>join a meeting next Monday at 10 of to chat about the project's progress and get your feedback. Your input is super important to us, and we think it will be instrumental in making this project a success. Lastly, I just want to say thanks for your continued support and education. Together, we've got the potential to achieve some amazing results, and I'm confident that our combined efforts will lead to a successful outcome. Looking forward to seeing you at the meeting!</p>
---	--

На жаль, але з використанням бібліотеки помилок стало більше, ніж було. Можливо, такий інструмент буде непоганим для використання у невеликих роботах.

2.3 DeepL

DeepL – це провідна система у галузі машинного перекладу, яка відома своїми алгоритмами перекладу. Вона використовує штучний інтелект для забезпечення високоякісного перекладу. Ця система була запущена у 2017 році і з того часу вона зарекомендувала себе як один з найкращих сервісів по перекладанню [20].

У своїх архітектурі ця система використовує нейронні мережі для перекладу, що забезпечує високоякісний переклад. Переклади виглядають природними та точними. У своєму сервісі вони мають багато мов, і щороку ця колекція поповнюється новими.

Вони використовують зручний і зрозумілий інтерфейс, який дозволяє переводити текст і документи. Цей сервіс можна використовувати для навчання, роботи і життя. Використання сервісу для перекладу: DeepL має переклад з української. Результати у таблиці 2.13.

Таблиця 2.13 — Результат перекладу DeepL

Вхідне повідомлення	Вихідне повідомлення
Зітхнувши важко, він усвідомив, що доля знову підклала йому свиню, змусивши пройти крізь терні до зірок, аби досягти омріяного успіху	Sighing heavily, he realized that fate had dealt him another pig, forcing him to go through the thorns to the stars to achieve his dream success.

Переклад достатньо точний, не має помилок і виглядає якісним, хоча і не знайшов аналогії для використання складних конструкцій.

Також DeepL можна використовувати для перефразування, переклавши текст спочатку на одну мову, а потім назад на оригінальну. Таким чином текст буде трохи видозмінюватися і виправляти помилки. Результат роботи у таблиці 2.14.

Таблиця 2.14 — Результат роботи перефразування з DeepL

Вхідне повідомлення	Вихідне повідомлення
Hey their, Hope you're doing well! I just wanted to shear some exciting news about our upcoming project. We've	Hi there, I hope you are doing well! I just wanted to share some exciting news about our upcoming project. We've

Продовження таблиці:

Вхідне повідомлення	Вихідне повідомлення
<p>made a lot of progress over the past few months, and I thought you'd like to here about it.</p> <p>First off, our team has wrapped up the initial faze of the project, which involved a lot of research and data colletion. We've gained some really valuable insights that will help guild our next steps. Now, we're moving into the developepment phase where we'll start implementing all the strategies we've come up with.</p> <p>Also, I'd love for you to join a meeting next Monday at 10 AM to chat about the project's progress and get your feedback. Your input is super inportant to us, and we think it will be instrumental in making this project a success.</p> <p>Lastly, I just want to say thanks for your continued support and dedication. Together, we've got the potential to achieve some amazing results, and I'm confident that our combined efforts will lead to a succsesful outcome.</p> <p>Looking forward to seeing you at the meeting!</p>	<p>made a lot of progress over the past few months and I thought you'd like to hear about it.</p> <p>First, our team has completed the initial phase of the project, which involved a lot of research and data collection. We received really valuable information that will help the guild take the next steps. Now we are moving to the development phase, where we will start implementing all the strategies we have developed.</p> <p>In addition, I would like you to join the meeting next Monday at 10am to chat about the progress of the project and get your feedback. Your input is very important to us and we believe it will help make this project a success.</p> <p>Finally, I just want to thank you for your continued support and dedication. Together we can achieve impressive results and I am confident that our joint efforts will lead to a successful outcome.</p> <p>I look forward to meeting you soon!</p>

Не дивлячись на те, що DeepL не може перефразувати повідомлення у заданому стильовому забарвленні, з базовою задачею він справляється і виправляє помилки в тексті.

2.4 Paraphrasing tools

На даний момент існує багато онлайн сервісів з перефразуванням, які можуть стати у нагоді при вирішенні питанні перефразування та стилістичного забарвлення. З них багато схожих, тому немає сенсу перевіряти усі на якість роботи. Вони користуються популярність у сферах навчання і роботи, де кожен добивається найбільш індивідуального результату.

Перефразування в офіційному стилі для таких сервісів не є проблемою. Результат можна побачити у таблиці 2.15.

Таблиця 2.15 — Результат перефразування повідомлення у офіційному стилі з онлайн сервісами

Вхідне повідомлення	Вихідне повідомлення
<p>Hey their,</p> <p>Hope you're doing well! I just wanted to shear some exciting news about our upcoming project. We've made a lot of progress over the past few months, and I thought you'd like to here about it.</p> <p>First off, our team has wrapped up the initial faze of the project, which involved a lot of research and data colletion. We've gained some really valuable insights that will help guild</p>	<p>Hey there.</p> <p>I hope you're doing well! I simply wanted to share some amazing developments regarding our next endeavor. We've made a lot of progress in the last several months, and I thought you'd be interested in hearing about it.</p> <p>First and foremost, our team has completed the initial phase of the project, which included extensive study and data collection. We've gathered</p>

Продовження таблиці:

Вхідне повідомлення	Вихідне повідомлення
<p>our next steps. Now, we're moving into the development phase where we'll start implementing all the strategies we've come up with.</p> <p>Also, I'd love for you to join a meeting next Monday at 10 AM to chat about the project's progress and get your feedback. Your input is super important to us, and we think it will be instrumental in making this project a success.</p> <p>Lastly, I just want to say thanks for your continued support and dedication. Together, we've got the potential to achieve some amazing results, and I'm confident that our combined efforts will lead to a successful outcome.</p> <p>Looking forward to seeing you at the meeting!</p>	<p>some extremely helpful insights that will help us plan our next actions. We are now entering the development phase, during which we will begin applying all of the plans we have devised.</p> <p>Also, I'd like to invite you to a meeting next Monday at 10 a.m. to discuss the project's development and receive your input. Your feedback is extremely valuable to us, and we believe it will be vital in making this project a success.</p> <p>Finally, I'd like to thank you for your continuing support and dedication. We have the ability to create incredible outcomes when we work together, and I am convinced that our combined efforts will be successful.</p> <p>I look forward to seeing you at the meeting.</p>

Результат роботи гарний. Такі онлайн сервіси мають у собі багато інструментів для роботи, які можуть знадобитися при роботі з соціальними мережами. Тому необхідно мати на увазі і цей інструмент при реалізації інтеграції перефразувачів у соціальні мережі.

2.5 Використані інструменти

При проведенні експерименту використовувалась мова програмування Python 3.9, яка відома гнучким синтаксисом і великою кількістю адаптованих бібліотек для штучного інтелекту, нейронних мереж та глибинного навчання.

2.5.1 Transformers

Transformers — це відома бібліотека з відкритим вихідним кодом, розроблена компанією Hugging Face. Вона надає велику кількість моделей-трансформерів для роботи, у яких використання у напрямках обробки природної мови майже не обмежене. Містить у собі велику кількість існуючих моделей та токенізаторів для використання. Тому користувачі можуть легко обирати інструменти для своїх робіт. Також компанія надає зручну документацію використання для користувачів. Також бібліотека містить інструменти для налаштування та навчання моделей.

Тому ця бібліотека користується великою кількістю запитів для роботи та розробки нових інструментів.

2.5.2 Pytorch

PyTorch — це потужний та гнучкий фреймворк для глибокого навчання, що розроблений компанією Facebook AI Research lab. Він став одним із найпопулярніших фреймворків для роботи з нейронними мережами: розробкою і навчанням. На відміну від попередників, цей фреймворк гнучкий, бо у своїй основі використовує графи.

Основними компонентами фреймворку є тензори. Це багатовимірні масиви, які схожі на NumPy масиви, але з підтримкою GPU процесорів для швидших обчислень.

PyTorch надає велику кількість інструментів для легкої побудови і навчання створених нейронних мереж, а його розширюваність дозволяє додавати чи змінювати створену мережу. Це надає зручність розробникові у виправленні та налаштуванні створеної моделі.

Завдяки цьому інструменту можна створювати системи для комп'ютерного зору, обробки природної мови, рекомендаційних систем та робототехніки.

2.6 Результати роботи

Під час дослідження інструментів для перефразування повідомлень, було оцінено декілька існуючих моделей штучного інтелекту та сервісів для роботи з такими задачами. Усі результати оцінювались мануально, без використання побічних інструментів. У кожного сервісу чи системи оцінювались також враховувались легкість в роботі, зрозумілість. Але такі результати не впливають саме на якість роботи. Тому у фінальну таблицю вони не занесені. Оцінювалися здатність перефразувати і виправляти помилки. Також перекладати текст. Тому потрібно підвести результати задля оцінки існуючих способів перефразування. Результати можна побачити у таблиці 2.16.

Таблиця 2.16 — Результати роботи моделей для перефразування

Методи оцінки	Моделі та сервіси для перефразування			
	BERT	GPT	T5	RoBERTa
Виправлення помилок	Якісне виправлення	Якісне виправлення	Неякісне виправлення	Неякісне виправлення

Продовження таблиці:

Зміна стилістичного тексту або перефразування	Якісне перефразування	Якісне перефразування	Неякісне перефразування	Неякісне перефразування
Переклад	Якісний переклад	Якісний переклад	Неякісний переклад	Неякісний переклад
API для використання (або інші методи для імплементації)	Має	Має	Має	Має

Методи оцінки	Моделі та сервіси для перефразування		
	TextBlob	DeepL	Paraphrasing tools
Виправлення помилок	Неякісне виправлення	Якісне виправлення	Якісне виправлення
Зміна стилістичного тексту або перефразування	Немає	Якісне перефразування	Немає
Переклад	Немає	Якісний переклад	Немає
API для використання (або інші методи для імплементації)	Має	Має	Має

Таким чином, абсолютними лідерами є генератори BERT та GPT, найбільші моделі для обробки природної мови. Вони зовсім не допустили помилок, а робота з ними проста і легко автоматизована. Не буде важким використання для одного разу, і так само легко зробити сервіс з API цих моделей для використання. Також вони вміють розпізнавати контекст, підлаштовувати текст для нової аудиторії, користуватися зображеннями та емодзі для покращення читабельності тексту. Також вони можуть генерувати контент у соціальних мережах для людей, створювати нові унікальні теми та підлаштовувати цей контент під читачів. Такі інструменти корисні для використання і для звичайних людей, і для компаній, тому що ці інструменти дозволяють швидко автоматизувати усю рутинну роботу, якої в соціальних мережах може бути достатньо, аналізувати аудиторію, її настрої та потреби для збільшення охоплення та створення нового креативного контенту. Також можна легко автоматизувати створення коментарів і відповідей на них, щоб підтримувати лояльність аудиторії.

T5 та RoBERTa. Вони для роботи потребують навчання та виправлення деяких моментів. Вони не мають API, тому використання обмежене у рамках існуючих моделей і наявності інструментів, що дозволять навчати та виправляти модель на власному пристрої. Це не найгірший варіант для використання, але ці моделі не гнучкі, не дуже зручні в користуванні. Для одноразового використання не буде найкращим інструментом.

DeerL і як перекладач, і як інструмент для перефразовування впорався із задачею. Для імплементації у соціальні мережі його можна використовувати як перекладач, хоча є більша кількість систем, які краще впораються з перефразовуванням. Точніше, в багатьох системах можна контролювати перефразовування.

Онлайн перефразувачі можуть гарно впоратися з багатьма задачами передачі та створення інформації у соціальних мережах. Вони гарно працюють, хоча не такі гнучкі, як інші системи.

ВИСНОВКИ

У ході дослідницької роботи було зроблено дослідження на тему перефразування повідомлень. Під час дослідження було виявлено проблеми, пов'язані з цією темою, та актуальність дослідження. Було зосереджено увагу на виявленні причин перефразування та аналізу систем, які можуть стати вирішенням цих проблем.

Протягом практики було зосереджено увагу на вивченні та аналізі сучасних технологій та алгоритмів машинного навчання, які застосовуються для генерації, перефразування та оптимізації текстових повідомлень в соціальних мережах. Особливий акцент було зроблено на дослідженні існуючих систем і можливість їх додавання до соціальних мереж.

Важливою частиною дослідження стало вивчення методів штучного інтелекту, таких як глибоке навчання та нейронні мережі, зокрема архітектури трансформерів, що демонструють високу ефективність у задачах обробки природної мови. Було проведено експерименти з різними моделями для визначення оптимальних підходів до перефразування тексту, що сприяло глибшому розумінню можливостей і обмежень існуючих технологій.

Під час роботи було зроблено великий аналіз існуючих систем, їх порівняння та можливості вирішення проблем перефразування. Було аналізовано результати і обрані найкращі системи для перефразувань.

Завершальним етапом дослідження стало підготовка та представлення звіту, який детально відражає всі аспекти проведеної роботи, включаючи огляд використаних методів аналіз отриманих результатів та висновки щодо перспектив застосування перефразування для повідомлень у соціальних мережах. Було виявлено потенційні проблеми, методи їх вирішення, покращення обміну інформації між користувачами соціальних мереж та створення унікального контенту.

Таким чином, ця робота може вплинути на розробку соціальних мереж, адже вона відкриває проблеми та додає рішення, які можуть значно вплинути на якість роботи у соціальних мережах для користувачів і компаній, для читачів і аудиторії.

Загальний вплив та суспільне сприйняття технологій штучного інтелекту, зокрема мереж побудованих по типу архітектуру трансформер, значно зросло протягом останніх років. Після виходу в світ GPT 3.5 в світі інформаційних технологій почався бум, який скоріш за все повністю трансформує як галузь в цілому так і соціальні мережі зокрема. Дана робота пропонує до уваги читачів цікаве порівняння того, які доступні мережі найкраще адаптовані до роботи в контексті соціальних мереж, зокрема в межах завдання по перефразуванню коротких повідомлень.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Social media. Wikipedia. URL: https://en.wikipedia.org/wiki/Social_media (дата звернення: 27.05.2024).
2. Nordquist R. Message (communication). ThoughtCo. URL: <https://www.thoughtco.com/message-communication-term-1691309> (дата звернення: 27.05.2024).
3. Wang Z., Liu Z. Analysis of Social Media Impact on Society. Journal of Information and Communication. URL: <https://ojs.ukscip.com/journals/jic/article/view/220> (дата звернення: 27.05.2024).
4. Classmates.com. Wikipedia. URL: <https://en.wikipedia.org/wiki/Classmates.com> (дата звернення: 27.05.2024).
5. Facebook. Wikipedia. URL: <https://en.wikipedia.org/wiki/Facebook> (дата звернення: 27.05.2024).
6. Кількість користувачів соцмереж в Україні за рік зменшилася на 10%. MediaSapiens. URL: <https://ms.detector.media/internet/post/34670/2024-04-15-kilkist-korystuvachiv-sotsmerezh-v-ukraini-za-rik-zmenshylasya-na-10/> (дата звернення: 27.05.2024).
7. Alalwan A. A., Rana N. P., Dwivedi Y. K. Social media in marketing: A review and analysis of the existing literature. Journal of Retailing and Consumer Services. URL: <https://www.sciencedirect.com/science/article/abs/pii/S0969698922001692> (дата звернення: 27.05.2024).
8. Wilson M. Most Popular Social Media Platforms in 2024. Shopify. URL: <https://www.shopify.com/blog/most-popular-social-media-platforms> (дата звернення: 27.05.2024).
9. Loudenback T. These are the most popular emojis in the US. Business Insider. URL: <https://www.businessinsider.com/most-popular-emoji-us-apple-2017-11?IR=T> (дата звернення: 27.05.2024).

10. Докінз Р. Егоїстичний ген. Київ: Наш Формат, 2016. 440 с.
11. Wang S., Li Q., Fan Y. Impact of social media use on mental health during the COVID-19 pandemic: A case study in China. *Digital Communications and Networks*. URL: <https://www.sciencedirect.com/science/article/pii/S2095809919304928> (дата звернення: 27.05.2024).
12. Natural Language Processing Applications. Data Science Dojo. URL: <https://datasciencedojo.com/blog/natural-language-processing-applications/> (дата звернення: 27.05.2024).
13. Chen X., Zhang M., Wang Y. The role of social media in crisis management during the COVID-19 pandemic. *Journal of Safety Science and Resilience*. URL: <https://www.sciencedirect.com/science/article/pii/S2666651022000146> (дата звернення: 27.05.2024).
14. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A. N., Kaiser Ł., Polosukhin I. Attention is all you need. In: *Advances in Neural Information Processing Systems*, 2017, vol. 30.
15. Devlin J., Chang M.-W., Lee K., Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2019, vol. 1, pp. 4171-4186.
16. Generative pre-trained transformer. Wikipedia. URL: https://en.wikipedia.org/wiki/Generative_pre-trained_transformer (дата звернення: 27.05.2024).
17. T5 (language model). Wikipedia. URL: [https://en.wikipedia.org/wiki/T5_\(language_model\)](https://en.wikipedia.org/wiki/T5_(language_model)) (дата звернення: 27.05.2024).
18. Raffel C., Shazeer N., Roberts A., Lee K., Narang S., Matena M., Zhou Y., Li W., Liu P. Exploring the Limits of Transfer Learning with a Unified Text-to-

- Text Transformer. arXiv. URL: <https://arxiv.org/abs/1907.11692> (дата звернення: 27.05.2024).
19. TextBlob: Simplified Text Processing. Read the Docs. URL: <https://textblob.readthedocs.io/en/dev/> (дата звернення: 27.05.2024).
20. Why DeepL? DeepL. URL: <https://www.deepl.com/whydeepl> (дата звернення: 27.05.2024).
21. Guo M., Ainslie J., Uthus D., Ontanon S., Ni J., Sung Y.-H., Yang Y. LongT5: Efficient Text-To-Text Transformer for Long Sequences.
22. Фігейра А., Ваз Б. Survey on Synthetic Data Generation, Evaluation Methods and GANs. Mathematics. 2022, т. 10, № 15, с. 2733.
23. Zhou W., Lee D.-H., Selvam R. K., Lee S., Lin B. Y., Ren X. Pre-training Text-to-Text Transformers for Concept-centric Common Sense.
24. Akbari H., Yuan L., Qian R., Chuang W.-H., Chang S.-F., Cui Y., Gong B. VATT: Transformers for Multimodal Self-Supervised Learning from Raw Video, Audio and Text. In: Advances in Neural Information Processing Systems 34 (NeurIPS 2021).
25. Jin M., Aletras N. Complaint Identification in Social Media with Transformer Networks.