

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

Національний університет кораблебудування імені адмірала Макарова

Навчально-науковий інститут комп'ютерних наук та управління проектами
Кафедра інформаційних управляючих систем та технологій

"Допущений до захисту"

Завідувач кафедри ІУСТ



к.т.н, доц. Михелєв І.Л.

" ____ " _____ 2024 р.

КВАЛІФІКАЦІЙНА РОБОТА

на здобуття ступеня вищої освіти "магістр"

на тему: Дослідження методів машинного навчання для узагальнення тексту
новин

Виконав: студент групи 6141 м



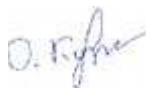
Кучеренко К.Р.

(підпис)

Керівник роботи:

доцент кафедри ІУСТ, к.т.н.,

(посада, науковий ступінь вчене звання)



Гайдаєнко О.В.

(підпис)

м. Миколаїв – 2024 р

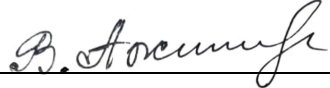
МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

Національний університет кораблебудування імені адмірала Макарова

Навчально-науковий інститут комп'ютерних наук та управління проектами
Кафедра інформаційних управляючих систем та технологій
Спеціальність 122 "Комп'ютерні науки"
Освітня програма "Комп'ютерні науки"

"ЗАТВЕРДЖУЮ"

Гарант освітньої програми



к.т.н, доц. Ажищев В.Ф.
" ____ " _____ 2024 р.

ЗАВДАННЯ НА КВАЛІФІКАЦІЙНУ РОБОТУ на здобуття ступеня вищої освіти "магістр"

Студенту: Кучеренко Кирило Руслановичу

1. Тема роботи: Дослідження методів машинного навчання для узагальнення тексту новин

Керівник роботи: Гайдаєнко Оксана Володимирівна

Затверджені наказом ректора № 1055-уч. від "08" жовтня 2024 року.

2. Термін подання роботи: 15 грудня 2024 р.

3. Вихідні дані до проекту (роботи) ДСТУ щодо обробки інформації, літературні джерела, технічна документація на існуючі аналоги систем, матеріали наукового стажування

4. Перелік питань, що належать до розробки (найменування розділів)

1. Аналіз сучасних підходів до узагальнення текстів новин.

2. Вибір та обґрунтування методів машинного навчання.

3. Проектування архітектури та розробка концепції веб-додатку.

5. Перелік презентаційних матеріалів

1. Постановка задачі, діаграма варіантів використання, концептуальна модель.

2. Огляд існуючих моделей, їх переваги та недоліки.

3. Архітектура системи, огляд алгоритму роботи моделі для узагальнення текстів новин.

4. Приклад роботи процесу узагальнення, метрики, результат.

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
1	Гайдаєнко О.В.	05.09.2023	22.12.2023
2	Гайдаєнко О.В.	23.12.2023	15.06.2024
3	Гайдаєнко О.В.	15.06.2024	25.09.2024
4	Гайдаєнко О.В.	25.09.2024	14.11.2024
5	Гайдаєнко О.В.	25.09.2024	14.11.2024
6	Гайдаєнко О.В.	25.09.2024	14.11.2024

7. Дата видачі завдання 05 вересня 2022 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів дипломного проекту (роботи)	Строк виконання етапів роботи	Примітка
1	Дослідження та аналіз предметної галузі	22.12.2024	
2	Розробка концепції	15.03.2024	
3	Розділ з розробки проектних рішень по системі (та її частинах)	15.06.2024	
4	Реалізація проекту та розробка робочої документації	25.09.2024	
5	Обчислення економічної ефективності	14.11.2024	
6	Вирішення питань охорони праці та навколишнього середовища	14.11.2024	
7	Оформлення пояснювальної записки	14.11.2024	
8	Подання магістерської роботи на перевірку на плагіат	16.11.2024	
9	Подання магістерської роботи рецензенту	20.11.2024	
10	Подання магістерської роботи на державний захист	15.12.2024	

Студент

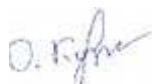


(підпис)

Кучеренко К.Р

(ПІБ)

Керівник роботи



(підпис)

Гайдаєнко О.В.

(ПІБ)

АНОТАЦІЯ

Система була оптимізована для зменшення обчислювальних витрат, що дозволяє її впровадження в реальні додатки, такі як новинні агрегатори, мобільні додатки або корпоративні інформаційні системи. Запропонований підхід демонструє У роботі розглянуто процес розробки системи для автоматизованого узагальнення новинних текстів на основі моделі T5. Зокрема, досліджено методи машинного навчання для забезпечення точного й смислово цілісного скорочення текстів. Основну увагу приділено аналізу сучасних технологій, розробці архітектури системи, її практичному впровадженню та оцінці результатів за допомогою метрик. Окрім технічних аспектів, у роботі висвітлено вплив системи на економічну ефективність, навколишнє середовище та умови праці, запропоновано способи мінімізації негативного впливу у використанні автоматизації текстових даних, підвищуючи ефективність роботи з інформацією та сприяючи сталому розвитку ІТ-індустрії.

Значна увага приділена налаштуванню моделі для роботи з багатомовними текстами, що розширює можливості використання системи в міжнародному контексті. Результати показали, що запропонований підхід може бути інтегрований у сервіси глобального рівня для забезпечення якісного оброблення інформації.

Ключові слова: машинне навчання, T5, ROUGE, BLEU, METEOR, BERTScore, API.

ABSTRACT

This study explores the development of a system for automated news text summarization based on the T5 (Text-to-Text Transfer Transformer) model. It investigates machine learning methods to ensure accurate and semantically coherent text compression. The focus is on analyzing modern technologies, designing the system architecture, implementing it in practice, and evaluating results using metrics such as ROUGE, BLEU, METEOR, and BERTScore. Beyond technical aspects, the study addresses the system's economic efficiency, environmental impact, and working conditions, proposing ways to minimize adverse effects.

The system has been optimized to reduce computational costs, enabling its application in real-world solutions such as news aggregators, mobile applications, or corporate information systems. This approach highlights the potential for automated text data processing to improve efficiency and contribute to the sustainable development of the IT industry.

Special emphasis has been placed on configuring the model to work with multilingual texts, extending the system's applicability in international contexts. The results demonstrate that the proposed approach can be integrated into global-level services to ensure high-quality information processing.

Additionally, the study examines the ethical implications of text-processing technologies, particularly concerning combating misinformation and data manipulation. The proposed system can serve as a valuable tool for media organizations and educational institutions.

Keywords: machine learning, T5, ROUGE, BLEU, METEOR, BERTScore, API.

ЗМІСТ

Перелік умовних позначень.....	10
Вступ.....	12
РОЗДІЛ 1. ДОСЛІДЖЕННЯ МЕТОДІВ МАШИННОГО НАВЧАННЯ ДЛЯ УЗАГАЛЬНЕННЯ ТЕКСТУ НОВИН ТА АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ.....	13
1.1 Огляд існуючих методів узагальнення текстів новин.....	13
1.2 Критичний аналіз переваг та недоліків існуючих підходів.....	15
1.3 Огляд практичних реалізацій та інструментів для узагальнення текстів новин.....	16
1.4 Виклики та обмеження у застосуванні методів машинного навчання для узагальнення новин.....	17
1.5 Перегляд сучасних досягнень у галузі машинного навчання.....	19
1.6 Вибір оптимальних методів для узагальнення новин.....	20
1.7 Аналіз можливостей та обмежень обраної моделі.....	22
1.8 Аналіз можливостей покращення вибраного методу.....	23
1.9 Концепція та постановка задачі.....	24
Висновок до розділу 1.....	25
РОЗДІЛ 2. ПРОЕКТУВАННЯ ІНФОРМАЦІЙНОЇ СИСТЕМИ.....	26
2.1 Побудова моделі варіантів використання.....	27
2.2 Побудова концептуальної моделі.....	27
2.3 Рішення з інформаційного забезпечення.....	28
2.4 Математична модель.....	30
2.5 Вибір мови проектування для додатку узагальнення текстів новин.....	30
2.6 Вибір і підготовка даних.....	32
2.6.1 Опис джерел даних (новинні статті, тематичні сайти, агрегатори новин).....	33
2.6.2 Процес збору даних: використання API чи веб-скрейпінгу.....	34

2.6.3 Попередня обробка тексту: очищення від HTML-тегів, видалення стоп-слів, токенізація, лематизація тощо.....	35
Висновок до розділу 2.....	45
РОЗДІЛ 3. РОЗРОБКА СИСТЕМИ.....	46
3.1 Розроба фізичної та логічної моделі, опис діаграми розгортання сиситеми.....	46
3.2 Аналіз прикладів отриманих узагальнень та їх відповідність оригінальним текстам.....	49
3.3 Інтеграція результатів у прикладне рішення.....	50
3.3.1 Розробка веб-інтерфейсу для доступу до системи узагальнення новин..	51
3.3.2 Демонстрація практичної роботи системи: введення тексту новин та отримання стислого змісту.....	52
3.3.3 Перспективи впровадження у реальних умовах, наприклад, у новинні агрегатори чи мобільні додатки.....	53
Висновок до розділу 3.....	55
РОЗДІЛ 4. РОЗРАХУНОК І ОЦІНКА ЕКОНОМІЧНОГО ЕФЕКТУ ВІД СИСТЕМИ УЗАГАЛЬНЕННЯ ТЕКСТУ НОВИН ЗА ДОПОМОГОЮ МЕТОДІВ МАШИННОГО НАВЧАННЯ.....	56
4.1 Загальна характеристика роботи.....	56
4.2 Розрахунок витрат на проектування та впровадження системи.....	58
4.3 Розрахунок економічної ефективності розробки і впровадження системи..	59
Висновок до розділу 4.....	60
РОЗДІЛ 5. ОХОРОНА ПРАЦІ.....	61
5.1 Вступ.....	61
5.2 Аналіз небезпечних і шкідливих факторів.....	63
5.2.1 Освітленість робочого місця.....	63
5.2.2 Вплив випромінювання.....	64
5.2.3 Вплив електричного струму.....	64
5.3 Розрахунок освітленості робочого місця.....	65

Висновок до розділу 5.....	65
 РОЗДІЛ 6. ОХОРОНА НАВКОЛИШНЬОГО СЕРЕДОВИЩА.....	66
6.1 Вступ.....	66
6.2 Загальні відомості про забруднення.....	67
6.3 Заходи щодо запобігання забруднення навколишнього середовища.....	68
Висновок до розділу 6.....	68
 Висновки.....	69
Список літератури.....	71
 Додаток А Технічне завдання.....	72
Додаток Б Інструкція користувача.....	75
Додаток В Код програми для узагальнення текстів новин.....	76

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

T5 (Text-to-Text Transfer Transformer) – трансформерна модель глибокого навчання, що перетворює один текстовий формат у інший для вирішення задач обробки природної мови.

ROUGE (Recall-Oriented Understudy for Gisting Evaluation) – метрика для оцінки якості текстового узагальнення, яка базується на порівнянні збігів фраз, слів або підрядків між референсним і отриманим текстами.

BLEU (Bilingual Evaluation Understudy) – метрика оцінки машинного перекладу або текстового узагальнення, яка визначає схожість між згенерованим текстом і референсним шляхом аналізу збігів n-грам.

METEOR (Metric for Evaluation of Translation with Explicit ORdering) – метрика оцінки текстів, яка враховує збіги на рівні слів, коренів, синонімів і словосполучень із референсним текстом.

BERTScore – метрика для оцінки текстового узагальнення, яка використовує семантичну подібність між референсним текстом і узагальненням на основі векторних подань слів.

API (Application Programming Interface) – інтерфейс програмування додатків, що дозволяє взаємодіяти з зовнішніми або внутрішніми компонентами програмного забезпечення.

ML (Machine Learning) – машинне навчання, підгалузь штучного інтелекту, яка використовує алгоритми для аналізу даних, побудови моделей і прийняття рішень.

NLP (Natural Language Processing) – обробка природної мови, галузь штучного інтелекту, яка займається аналізом і генерацією текстів.

TF-IDF (Term Frequency-Inverse Document Frequency) – статистичний метод для оцінки важливості термінів у текстах на основі частоти їхнього використання.

JSON (JavaScript Object Notation) – текстовий формат для зберігання та передачі даних у вигляді структурованих об'єктів.

GPU (Graphics Processing Unit) – графічний процесор, що використовується для

пришвидшення обчислень у задачах глибокого навчання.

Hugging Face – платформа та бібліотека для роботи з моделями обробки природної мови, включно з T5.

Fine-tuning – процес донавчання попередньо тренованої моделі на конкретному наборі даних для досягнення кращих результатів у специфічних задачах.

Transformer – архітектура нейронної мережі, що використовується для обробки послідовностей даних, наприклад, текстів, завдяки механізму уваги.

Summarization – процес узагальнення тексту для виділення основного змісту, зменшуючи обсяг вихідного тексту.

ВСТУП

У сучасному світі зростає потреба в ефективних інструментах для обробки та аналізу текстових даних, оскільки інформаційні потоки стають дедалі ширшими. Особливо актуальною є автоматизація процедур реферування новин, оскільки вона дозволяє користувачам швидко виокремлювати важливі деталі з великих текстів. Збереження точності, релевантності та семантичної цілісності при створенні скорочених версій текстів ускладнює це завдання.

У цій статті розглядаються стратегії вирішення цієї проблеми за допомогою сучасних методів машинного навчання, а саме застосування моделі T5 (Text-to-Text Transfer Transformer). Цей метод поєднує в собі відмінну точність з адаптивністю до багатьох ситуацій використання, створюючи нові можливості для автоматизованої обробки тексту.

Метою роботи є створення методу реферування новинних статей, який гарантує як відмінні результати, так і доцільність його впровадження з економічної точки зору. Особливу увагу приділено налаштуванню процедур обробки тексту, оптимізації моделі та інтеграції системи в реальний робочий контекст. Аналіз впливу системи на навколишнє середовище, економічна ефективність і здатність забезпечити належні умови праці - все це має вирішальне значення.

Дослідження складається з кількох важливих етапів, серед яких вивчення останніх досягнень у галузі машинного навчання, створення архітектури системи, її фактичне використання та оцінка якості вихідних даних. Ефективність системи оцінюється за допомогою різних показників, зокрема ROUGE, BLEU, METEOR і BERTScore, які дають об'єктивну оцінку якості текстів, що створюються.

Оскільки розробка і застосування таких систем мають значний вплив на навколишнє середовище і здоров'я працівників, питанням охорони праці та навколишнього середовища приділяється особлива увага. Для того, щоб зменшити негативні наслідки використання ІТ-технологій, у дослідженні пропонуються конкретні заходи.

Основні результати кваліфікаційної роботи було апробовано на IX всеукраїнської науково-практичної конференції "Перспективні напрямки сучасної електроніки, інформаційних і комп'ютерних систем" яка проходила 27-29 листопада 2024 року [12].

РОЗДІЛ 1. ДОСЛІДЖЕННЯ МЕТОДІВ МАШИННОГО НАВЧАННЯ ДЛЯ УЗАГАЛЬНЕННЯ ТЕКСТУ НОВИН ТА АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ

1.1 Огляд існуючих методів узагальнення текстів новин

Моделі Seq2Seq. Моделі sequence-to-sequence (Seq2Seq) широко використовуються для автоматичного узагальнення текстів новин. Вони перетворюють вхідну послідовність слів (оригінальний текст) у вихідну послідовність (стислий текст), застосовуючи рекурентні або трансформерні нейронні мережі. Такий підхід дозволяє враховувати контекст і залежності між словами, що суттєво покращує якість узагальнення. Однак, ці моделі можуть стикатися з проблемою "забування" довгих залежностей у тексті, що іноді впливає на точність результату.

Методи з абстракційною генерацією. Абстракційна генерація полягає у створенні тексту, який не просто копіює частини вихідного матеріалу, а надає загальне абстрактне уявлення про його зміст. Цей підхід дозволяє зберегти ключову інформацію, створюючи більш осмислені узагальнення. Попри це, визначення коректного абстрактного опису може бути складним завданням, що іноді призводить до неточностей у кінцевому результаті.

Методи з використанням reinforcement learning. Reinforcement learning (підкріплювальне навчання) знаходить застосування в оптимізації моделей узагальнення шляхом ітеративного навчання. Моделі оцінюють результати своїх дій і постійно вдосконалюються. Це дозволяє досягти високої якості узагальнення, проте процес тренування потребує значних обчислювальних ресурсів і займає багато часу [1].

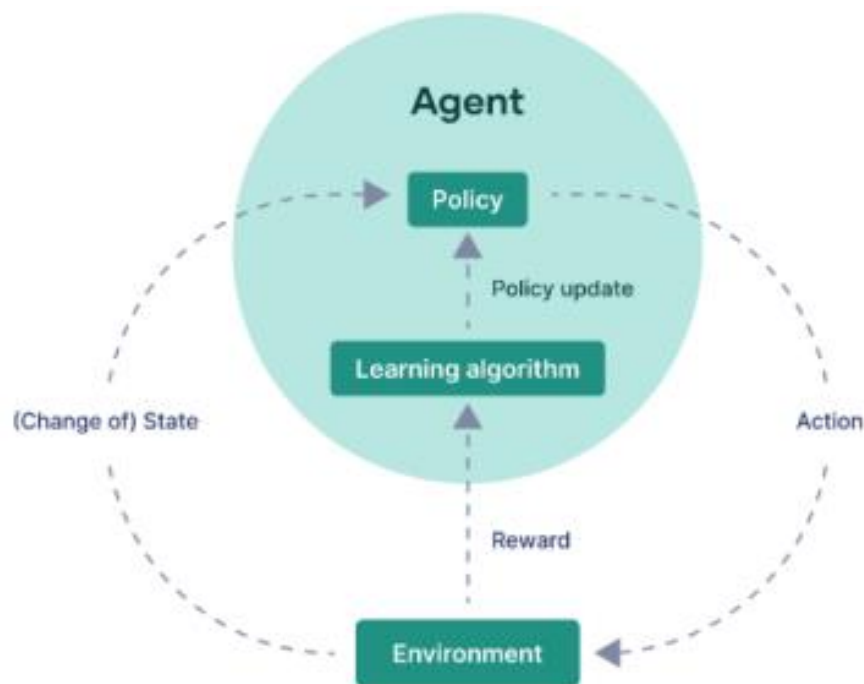


Рисунок 1 - Діаграма роботи алгоритму reinforcement learning.

Ансамблеві методи. Ансамблеві підходи покращують узагальнення тексту, об'єднуючи результати роботи кількох моделей. Щоб компенсувати недоліки окремих моделей, вони використовують різні структури або стратегії. Це особливо добре працює, коли продуктивність окремих моделей обмежена їхніми унікальними характеристиками.

Перевагою ансамблевих підходів є їхня здатність покращувати загальні результати завдяки інтеграції різних методів обробки тексту та джерел даних. Поєднання декількох моделей призводить до більш точного узагальнення і зменшує ймовірність упущення важливих деталей.

Однак навчання моделей і забезпечення інфраструктури для агрегування результатів ансамблевих підходів вимагають значних ресурсів. При використанні цього методу слід враховувати, що він може призвести до збільшення часу виконання і більшої потреби в складному апаратному забезпеченні.

Використання pre-trained моделей. Для завдань з узагальнення тексту часто використовують попередньо навчені моделі, такі як BERT або GPT. Ці моделі вже мають загальну інформацію, яку можна застосувати до конкретних завдань, оскільки вони були навчені на великих масивах даних. Навіть для

невеликих наборів даних їхня адаптація (тонка настройка) дає змогу зробити якісне узагальнення.

Перевага цього методу полягає в тому, що він використовує знання, отримані під час попереднього навчання, що значно скорочує час навчання і підвищує якість узагальнення. Попередньо натреновані моделі зберігають ключові елементи тексту і добре справляються із завданнями, що вимагають глибокого розуміння контексту.

Проте існують певні обмеження: моделі можуть не досягти необхідного рівня якості в умовах, коли бракує вихідних даних для тонкого налаштування.

1.2 Критичний аналіз переваг та недоліків існуючих підходів

Використання методів машинного навчання для узагальнення новинних історій - складна, але перспективна справа. Ця стратегія дає змогу автоматизувати обробку інформації, що є надзвичайно важливим з огляду на експоненціальне зростання обсягів даних. Але окрім переваг, існує низка проблем, які необхідно враховувати для ефективного використання цих методів. Для того, щоб збалансувати продуктивність і точність узагальнення, крім вибору відповідних алгоритмів, важливо гарантувати якість навчальних даних.

Методи машинного навчання мають кілька важливих переваг. По-перше, вони прискорюють обробку новин і підвищують продуктивність, автоматизуючи процес узагальнення. По-друге, навіть при великому обсязі даних їхня масштабованість гарантує високу ефективність. Здатність алгоритмів автоматично виявляти складні закономірності в текстах покращує якість узагальнення та підвищує об'єктивність результатів. Здатність алгоритмів адаптуватися до змін - ще одна суттєва перевага: навіть у динамічному новинному середовищі вони можуть вчитися на свіжих даних і залишатися актуальними.

Однак є й певні недоліки. Деякі підходи можуть призвести до надмірно спрощених або помилкових узагальнень, оскільки вони не враховують належним чином контекст. Додаткові труднощі виникають через високу непередбачуваність, неоднозначність і вразливість новинних текстів до впливу

навчальних даних. Крім того, для якісного навчання моделі часто потрібен великий обсяг даних, який не завжди можливо отримати. На об'єктивність узагальнень також впливає проблема упередженості навчальних даних. Моделям може бути важко встигати за швидкими змінами в новинах, оскільки вони повинні часто оновлюватися, щоб залишатися актуальними. Тому, незважаючи на величезний потенціал машинного навчання, його застосування вимагає обережної стратегії, яка враховує всі труднощі.

1.3 Огляд практичних реалізацій та інструментів для узагальнення текстів новин

Використання попередньо навчених моделей машинного навчання, таких як BERT, GPT, T5 або їхніх варіацій, є важливим компонентом сучасних систем узагальнення тексту. Генерування узагальнень - одне з завдань обробки природної мови, яке ці моделі дозволяють ефективно виконувати. Наприклад, завдяки своїй універсальній текстово-орієнтованій архітектурі модель T5 (Transformer of Text-to-Text Transfer Transformer) демонструє хорошу якість і спеціалізується на перетворенні тексту, що включає узагальнення. Для автоматичного створення коротких, змістовних речень з урахуванням контексту також добре зарекомендувала себе модель GPT від OpenAI.

Hugging Face Transformers з його зручним інтерфейсом і великою кількістю готових моделей став стандартом для реалізації узагальнень. Hugging Face підтримує численні складні моделі, які легко включати в серверні системи або веб-додатки. Можливості цієї бібліотеки роблять моделі дуже адаптивними, дозволяючи налаштовувати їх для певних застосувань, зокрема для узагальнення новин. Крім того, Hugging Face пропонує інтеграцію з різними екосистемами, такими як PyTorch та TensorFlow, що розширює можливості розробників.

Крім того, існують спеціалізовані програми для обробки тексту, такі як OpenAI API та Google Cloud Natural Language API. Вони надають готові до використання сервіси, які виконують узагальнення тексту, не вимагаючи розгортання локальних моделей. Ці сервіси пропонують масштабованим

додаткам високу доступність і точність, але їх використання може бути пов'язане зі значними витратами та ліцензійними обмеженнями.

Проблема вбудовування моделей узагальнення в реальні робочі процеси також береться до уваги в практичних реалізаціях. Наприклад, додатки для роботи з новинами, такі як Flipboard і Pocket, надають короткі резюме статей, використовуючи методи автоматичного стиснення тексту. У таких ситуаціях дуже важливо враховувати продуктивність системи і швидкість обробки на додаток до якості резюме. Тому для успішної реалізації програми для узагальнення новин вирішальне значення мають обрані моделі та технології.

1.4 Виклики та обмеження у застосуванні методів машинного навчання для узагальнення новин

Динамічний характер інформації є однією з головних перешкод для використання методів машинного навчання для узагальнення новин. Роль моделей ускладнюється тим, що новини представляють події, які часто змінюються, особливо якщо моделі навчаються на застарілих даних. Наприклад, модель може бути не знайома з новою термінологією, назвами організацій або подій, що знижує точність узагальнення. Це зумовлює необхідність частого перенавчання моделі та оновлення навчальних даних, що вимагає значних ресурсів.

Різноманітність стилів і структур новинних текстів - це ще одне питання. У новинних повідомленнях можна знайти різні стилі написання, рівні деталізації та форми відображення інформації. Це ускладнює завдання для моделей, оскільки вони можуть не враховувати значення або контекст деяких деталей, які є важливими для точного узагальнення. Наприклад, через те, що моделі не можуть правильно зрозуміти структуру тексту, вони можуть переоцінити дрібні деталі або прогавити важливі.

Іншим серйозним недоліком є проблема упередженості даних. Модель може давати упереджені узагальнення, якщо навчальні дані містять упередження, що стосуються, скажімо, соціальних чи політичних питань. Це викликає занепокоєння щодо об'єктивності та коректності результатів, особливо

при обговоренні делікатних тем. Щоб зменшити цей вплив, навчальні дані мають бути ретельно перевірені та відфільтровані, що робить процедуру більш складною.

І останнє, але не менш важливе: забезпечення продуктивності та масштабованості моделей є великою перешкодою. Обробка величезних обсягів даних у режимі реального часу необхідна для узагальнення текстів новин, особливо для комерційних додатків, які обслуговують тисячі клієнтів. Високі вимоги до обчислювальних ресурсів методів машинного навчання, особливо великих трансформаційних моделей, можуть підвищити вартість розробки та обслуговування системи. Як наслідок, розробники змушені шукати компроміси між швидкістю та якістю узагальнення.

1.5 Перегляд сучасних досягнень у галузі машинного навчання

- Моделі глибокого навчання. Використання глибоких нейронних мереж для рішення різноманітних завдань від розпізнавання об'єктів до машинного перекладу.
- Масштабність та обчислювальна потужність. Розвиток архітектур та алгоритмів, що сприяють ефективнішому використанню великих обчислювальних ресурсів.
- Застосування в інших галузях. Зростання застосувань машинного навчання в медицині, фінансах, автономних системах, промисловості, сільському господарстві та інших секторах.
- Автоматизоване машинне навчання. Зростання інструментів і платформ для автоматизованого створення, навчання і оптимізації моделей машинного навчання без глибокого розуміння користувача.
- Обробка природної мови (NLP). Покращення алгоритмів для розуміння та генерації природної мови, що призводить до розробки більш ефективних систем обробки мови [2].
- Зростання уваги до етичних питань. Розвиток методів для забезпечення справедливості, етичності та відповідального використання технологій машинного навчання.

1.6 Вибір оптимальних методів для узагальнення новин

Найбільш підходящою моделлю для виконання задачі узагальнення текстів новин, буде модель типу Seq2Seq (sequence-to-sequence). В минулому розділі був проведений аналіз та порівняння переваг та недоліків усіх можливих типів моделей, які можуть вирішувати дану задачу. Відповідно, ми можемо підвести підсумок, які саме переваги мають Seq2Seq моделі в порівнянні з іншими.

- Обробка послідовностей різної довжини: Моделі Seq2Seq призначені для роботи з послідовностями різної довжини на вході та виході. Це особливо важливо для завдань, де довжина вхідних та вихідних послідовностей може варіюватися, наприклад, у завданнях машинного перекладу або генерації тексту.

- Машинний переклад: Seq2Seq широко використовується в задачах машинного перекладу. Модель може вивчати відповідність між різними мовами та генерувати вихідний текст для будь-якої вхідної послідовності.

- Генерація тексту: Seq2Seq також використовується для генерації тексту. Наприклад, використовуючи вхідну послідовність слів або символів, модель може генерувати відповідну послідовність тексту.

- Використання у завданнях з рекурентними мережами: Seq2Seq може включати в себе рекурентні нейронні мережі (RNN) або більш ефективні архітектури, такі як Long Short-Term Memory (LSTM) або Gated Recurrent Unit (GRU) [3, 4, 5]. Це дозволяє моделі враховувати контекст та взаємодіяти зі вхідними даними в режимі послідовності.

- Застосування в задачах завдань узагальнення: Seq2Seq може використовуватися в різних задачах узагальнення, таких як аудіо переклад, генерація підписів до зображень, відповідь на запитання та інші. Модель дозволяє ефективно вирішувати завдання, де важлива взаємодія та залежність між елементами послідовності.

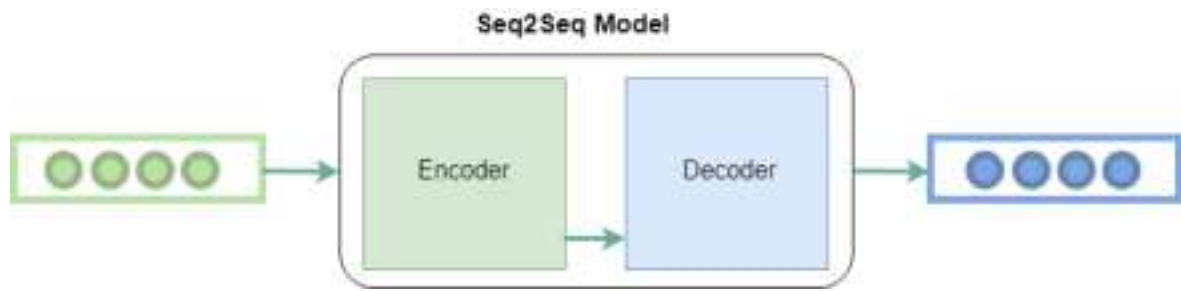


Рисунок 2 - Діаграма архітектури Seq2Seq моделі.

Seq2Seq, як випливає з назви, приймає на вхід послідовність слів (речення або речення) і виводить нову послідовність слів. Для цього він використовує рекурентну нейронну мережу (RNN). LSTM або GRU, дві більш складні версії RNN, використовуються замість базової версії, яка майже ніколи не використовується. Це пов'язано з проблемою зникаючого градієнта у RNN. Запропонована Google версія використовує LSTM. Кожного моменту часу для побудови контексту слова потрібні два вхідні дані. Слово "рекурентний" означає, що один з них надходить від користувача, а інший - з попереднього виходу (вихід слугує входом). Для реалізації кодера і декодера зазвичай використовуються трансформатори або рекурентні нейронні мережі (RNN).

Стек кодерів використовує глибокі шари нейронної мережі і перетворює вхідні слова на відповідні приховані вектори. Кожен вектор представляє поточне слово та його контекст. Кодер приймає вхідну послідовність, один лексему за раз, і використовує RNN або трансформатор для оновлення свого прихованого стану, який підсумовує інформацію у вхідній послідовності. Остаточний прихований стан кодера передається декодеру як вектор контексту.

Стек декодерів схожий на кодер. Він приймає на вхід прихований вектор, згенерований кодувальником, власні приховані стани та поточне слово, щоб створити наступний прихований вектор і, нарешті, передбачити наступне слово. Декодер використовує вектор контексту і початковий прихований стан для генерації вихідної послідовності, по одному токenu за

раз. На кожному кроці декодер використовує поточний прихований стан, вектор контексту і попередній вихідний лексему для генерації розподілу ймовірностей серед можливих наступних лексем. Токен з найбільшою ймовірністю обирається як вихідний, і процес продовжується доти, доки не буде досягнуто кінця вихідної послідовності.

1.7 Аналіз можливостей та обмежень обраної моделі

Моделі Seq2Seq (sequence-to-sequence) мають свої можливості та обмеження. Давайте розглянемо їх детальніше:

Можливості:

- Обробка послідовностей різної довжини: Можливість вирішення завдань машинного перекладу: Seq2Seq може ефективно вирішувати завдання машинного перекладу, де довжина вхідних та вихідних послідовностей може варіюватися.
- Генерація тексту: Генерація послідовностей: Модель може бути використана для генерації тексту, що корисно у завданнях, таких як автогенерація тексту або генерація підписів до зображень.
- Застосування у завданнях з рекурентними мережами: Врахування контексту: Seq2Seq, зокрема з рекурентними архітектурами, може добре враховувати контекст та взаємодіяти зі вхідними даними в режимі послідовності.
- Універсальність застосування: Широкий спектр задач: Модель може бути успішно використана для різних задач, таких як машинний переклад, генерація тексту, синтез мови тощо.

Обмеження:

- Проблема зниклих та вибухаючих градієнтів: Проблеми з тривалими послідовностями: При тривалих послідовностях можуть виникати проблеми з градієнтами, такі як зниклі та вибухаючі градієнти, що може призводити до погіршення якості моделі.
- Проблема вивчення контексту: Обмежена здатність враховувати довгострокові залежності: Деякі рекурентні архітектури можуть мати обмежену

здатність враховувати довгострокові залежності у тексті, що може бути проблемою для деяких задач.

- Залежність від якості даних: Потреба у великій кількості даних: Моделі Seq2Seq, особливо глибокі та складні, можуть вимагати великої кількості даних для успішного навчання та уникнення перенавчання.

- Обмеження архітектурою: Невластивість для деяких завдань: В деяких випадках, наприклад, в задачах, де важливий паралельний аналіз, інші архітектури, такі як трансформери, можуть бути більш ефективними.

- Вимоги до ресурсів: Велика обчислювальна потреба: Деякі Seq2Seq моделі, особливо з великою кількістю параметрів, можуть бути вимогливими до обчислювальних ресурсів для тренування та використання.

1.8 Аналіз можливостей покращення вибраного методу

На даний час, існуючі моделі машинного навчання, які використовуються для узагальнення текстів, навчені на не на специфічних датасетах, а точніше не на текстах новин. Через це, при використанні моделі, точність результатів узагальнення буде нижчою ніж у випадку, коли модель натренована на текстах новин.

В результаті проведеного аналізу, ми можемо вирішити цю проблему за допомогою такого методу як «Fine-tuning». Це процес внесення невеликих змін для отримання бажаного результату або продуктивності. При глибокому навчанні інший алгоритм глибокого навчання з тієї ж області програмується з використанням вагових коефіцієнтів навченої нейронної мережі. У нейронній мережі ваги встановлюють зв'язки між кожним нейроном одного шару та нейронами інших шарів. Точне налаштування використовується для подолання короткого набору даних та/або прискорення навчання, оскільки він вже має необхідні знання з попереднього алгоритму глибокого навчання.

Як результат, буде вирішена проблема з точністю генерації результатів. Самі узагальнення будуть більш підходящими для новин, так як буде використаний метод «Fine-tuning» для додаткового навчання моделі саме на текстах новин.

1.9 Концепція та постановка задачі

Основна мета цього дослідження — розробка концептуальної моделі додатку для автоматичного узагальнення текстів новин, зосереджуючись на вивченні та аналізі існуючих методів машинного навчання. Завдання полягає в тому, щоб визначити ефективні підходи до обробки тексту, такі як Seq2Seq моделі, абстракційна генерація, reinforcement learning, ансамблеві методи та використання попередньо навчених моделей. Кожен з цих підходів має свої переваги та недоліки, які слід враховувати при виборі оптимального методу.

Для досягнення мети дослідження необхідно вирішити низку важливих завдань. По-перше, необхідно зібрати та попередньо обробити текст новин, використовуючи методи лематизації, токенізації та видалення стоп-слів. Друга мета полягає у створенні та застосуванні моделей машинного навчання, які можуть надавати стислі, освітні резюме текстів. По-третє, дослідити і застосувати на практиці методи оцінки якості узагальнення, щоб ще більше вдосконалити моделі.

Розробка такого додатку вимагає вирішення кількох технічних питань, таких як вибір мови програмування та інструментів реалізації. Python було обрано як основну мову через його розгалужену екосистему нейромереж (TensorFlow, PyTorch) та модулів обробки природної мови (NLTK, SpaCy). Ми хочемо використовувати такі бази даних, як PostgreSQL або MongoDB для управління та зберігання даних, а також інтеграцію з API для отримання найсвіжіших новинних даних.

Таким чином, місія передбачає не лише дослідження існуючих підходів, але й розробку працездатної системи, яка пропонує високоточне та продуктивне автоматичне узагальнення новин. Метою дослідження є оптимізація процесу обробки тексту, підвищення якості резюме та розробка інструменту, який може принести користь як приватним особам, так і компаніям, що працюють з великими обсягами текстових даних.

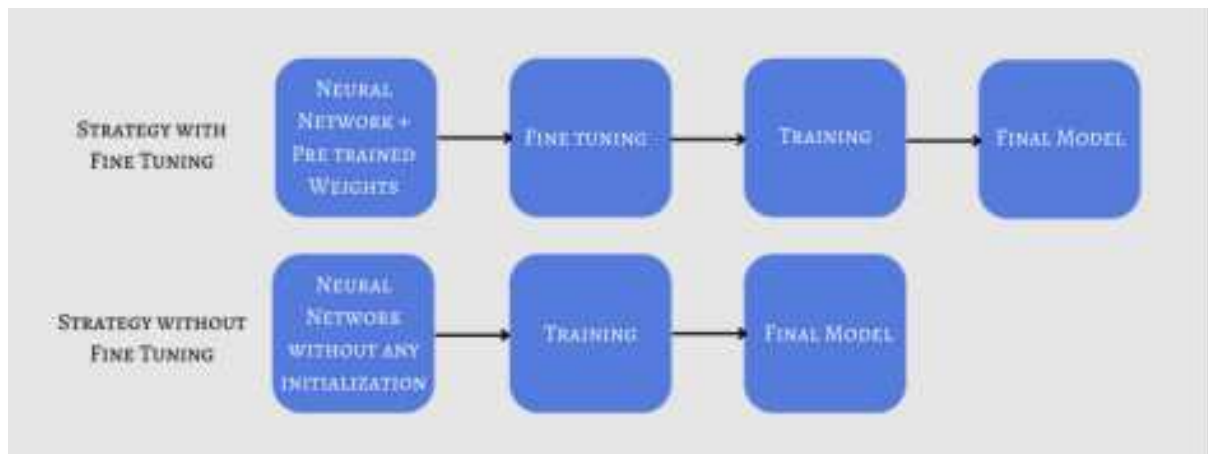


Рисунок 3 - Концепція роботи моделі за допомогою методу fine-tuning та без

Для того, щоб користувачу було зручно використовувати кінцевий функціонал, в майбутньому, може бути розроблений сервіс для узагальнення текстів новин. Сервіс буде включати в себе Front-end та Back-end частини, тобто буде розроблений повноцінний інтерфейс користувача, завдяки якому будь-хто зможе скористатись та перевірити функціонал та результати роботи.

Висновок до розділу 1

Оскільки методи машинного навчання, особливо моделі Seq2Seq, можуть обробляти послідовності різної довжини та враховувати контекст, вони досить успішно справляються з завданнями з узагальнення тексту. Ці моделі можна використовувати для вирішення різноманітних завдань, таких як генерація тексту, машинний переклад та інші складні завдання з обробки природної мови. Моделі Seq2Seq є універсальним і потужним інструментом для обробки новинного контенту завдяки їхній гнучкості в поєднанні з сучасними конструкціями, такими як трансформери.

Моделі Seq2Seq мають недоліки, такі як залежність від калібру і кількості навчальних даних, а також вимоги до обробки, незважаючи на їхні переваги, які включають високу точність і адаптивність. Використання складних архітектур, таких як LSTM або GRU, необхідне для вирішення проблем з вивченням довгострокових залежностей, а також потенційних труднощів з градієнтним

затуханням або вибухом. Оптимізація моделей для певних завдань є ще одним важливим компонентом, який дозволяє системі бути більш точною та гнучкою.

Одним із важливих способів підвищити точність узагальнення є використання стратегії тонкого налаштування для навчання моделі на новинних статтях. Ми можемо отримати більш релевантні результати, застосовуючи знання з раніше навчених моделей і налаштовуючи їх відповідно до певних вимог. Таким чином, ми зможемо розробити систему, яка вироблятиме чудові узагальнення, ідеально пристосовані до матеріалів новин.

Загалом, розробка додатку для узагальнення новин з використанням Seq2Seq моделей є перспективним напрямком, який відповідає сучасним вимогам ринку та потребам користувачів. Інтеграція таких методів із сучасними інструментами обробки природної мови та побудова інтуїтивно зрозумілого інтерфейсу користувача дозволить отримати ефективну та зручну систему автоматичного узагальнення новин. Це стане основою для створення сервісу, здатного обробляти великі обсяги даних і надавати користувачам швидкий доступ до необхідної інформації.

РОЗДІЛ 2. ПРОЕКТУВАННЯ ІНФОРМАЦІЙНОЇ СИСТЕМИ

2.1 Побудова моделі варіантів використання

Побудова моделі варіантів використання для додатку узагальнення текстів новин може бути цікавим завданням. Провівши аналіз, ми можемо виділити кілька можливих варіантів використання:

- Автоматичне узагальнення новинних статей: Додаток може автоматично узагальнювати довгі новинні статті в короткі зведення, щоб користувачі могли швидко отримати основну інформацію без необхідності читати всю статтю.
- Створення зведених новинних лент: Додаток може створювати персоналізовані ленти новин, узагальнюючи новини з різних джерел за вибором користувача.
- Система моніторингу новин: Додаток може відстежувати певні теми або ключові слова в новинах і автоматично узагальнювати новини на ці теми для користувача.
- Аналіз настроїв у медіа: Додаток може аналізувати новинні статті для виявлення настроїв (позитивних, негативних, об'єктивних) та створювати зведені звіти про настрої у медіа.
- Навчання машинного навчання: Використання додатком алгоритмів машинного навчання для автоматичного вдосконалення узагальнення новин і врахування виправлень, зроблених користувачами.

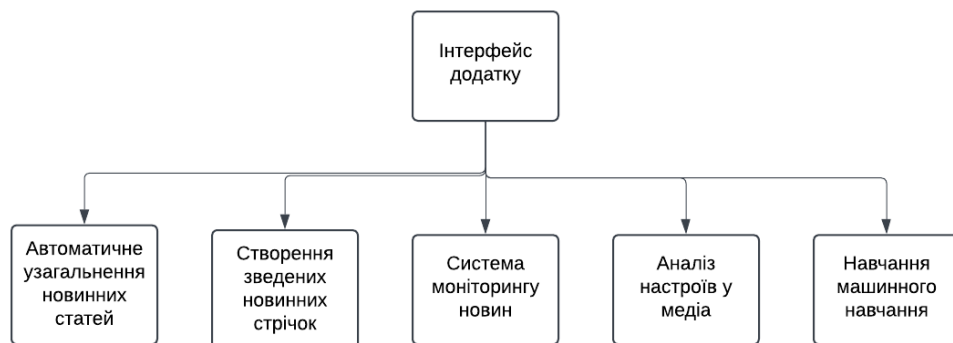


Рисунок 2.1 - Схематична модель

2.2 Побудова концептуальної моделі

Обробка вхідних даних - це перший крок у процесі узагальнення новин. На цьому етапі до програми надсилається або одна новина, або колекція новин, які потребують узагальнення. Попередня обробка цих текстів включає лематизацію, видалення стоп-слів, токенизацію та, за необхідності, нормалізацію тексту. Оскільки попередня обробка дозволяє нам підкреслити важливі частини тексту і видалити зайву інформацію, яка може спотворити результати, вона має вирішальне значення для забезпечення правильної роботи алгоритмів узагальнення [6].

Процес конспектування, розбитий на кілька важливих частин, є основним етапом роботи моделі. Алгоритм спочатку визначає, які речення або ключові фрази в тексті найбільш релевантні для резюме. Для цього можуть використовуватися методи машинного навчання, статистичний аналіз тексту або поєднання цих методів. Потім створюється стислий, але глибокий конспект тексту з використанням відібраних важливих термінів. Цей підхід є більш гнучким і адаптивним до вимог різних сценаріїв використання, оскільки користувач може змінювати параметри резюме, наприклад, довжину тексту або значення окремих фраз.

Результати та інші функції розглядаються на останньому кроці концептуальної моделі. Користувач може вказати текст, який потрібно обробити, і переглянути результати за допомогою інтуїтивно зрозумілого інтерфейсу, який надає узагальнений текст. Крім того, програма може включати можливість зберігати історію узагальнень, що буде корисно для подальшого пошуку результатів. Моніторинг якості є ключовим компонентом стратегії і може застосовуватися з використанням метрик оцінки тексту, таких як BLEU або ROUGE. Ця система гарантує якісне узагальнення, дозволяючи розробникам відстежувати продуктивність додатку та вносити постійні вдосконалення [7].

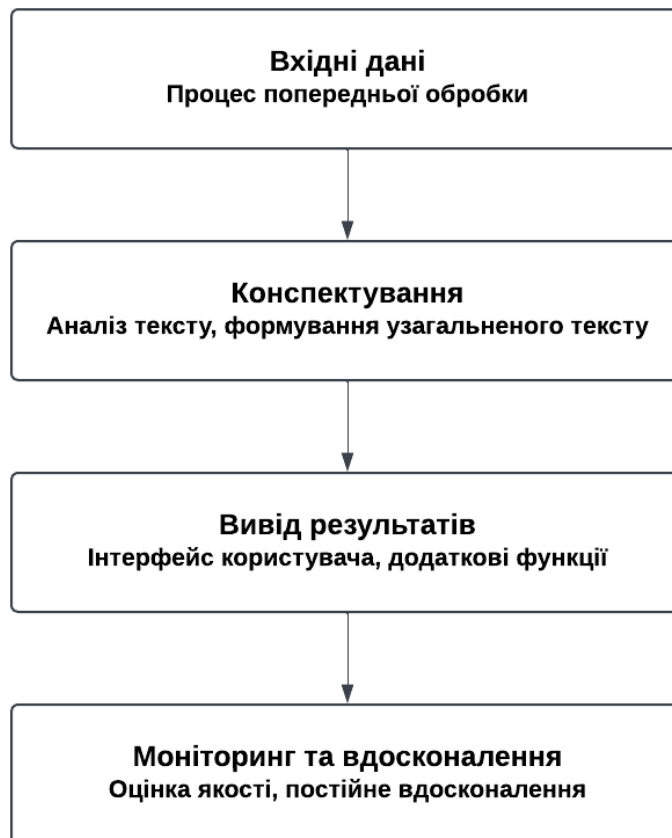


Рисунок 2.2 - Діаграма концептуальної моделі

2.3 Рішення з інформаційного забезпечення

Для того, щоб гарантувати функціональність програми зведення новин, необхідно налаштувати ефективну процедуру збору та обробки даних. Для збору текстів новин можна використовувати веб-скрепінг або API джерел новин, що дає змогу автоматично отримувати доступ до найсвіжішої інформації. Спеціалізовані бібліотеки, такі як NLTK або SpaCy, використовуються для обробки зібраних даних, щоб нормалізувати текст, усунути стоп-слова, токенізувати його та виконати інші необхідні завдання. Таким чином, текст стає готовим для подальшої обробки та аналізу моделями машинного навчання. На первинному етапі використовуються сучасні трансформантні моделі, такі як BERT або GPT, які завдяки своїм архітектурним можливостям пропонують чудові узагальнення, враховуючи при цьому структуру та контекст тексту [8].

– Зберігання, доступ і масштабованість. Зберігання даних є важливим аспектом інформаційного забезпечення додатку. Для цього можуть

використовуватися реляційні бази даних, такі як PostgreSQL, або документоорієнтовані системи, наприклад MongoDB, які дозволяють зберігати узагальнені тексти новин, дані користувачів і налаштування. Розробка веб- або мобільного інтерфейсу є ключовою для забезпечення зручного доступу користувачів до функцій додатку. Інтерфейс дозволяє вводити текст новин, переглядати узагальнені версії та налаштовувати параметри обробки. Особливу увагу слід приділити масштабованості додатку, щоб він міг обробляти великий обсяг даних і підтримувати велику кількість користувачів без зниження продуктивності.

— Безпека, аналітика та поліпшення. Захист даних користувачів є пріоритетним завданням. Для цього необхідно впроваджувати методи шифрування, багаторівневу аутентифікацію та інші заходи безпеки для захисту конфіденційної інформації. Інтеграція функціоналу аналітики дозволяє відстежувати якість узагальнень, використовуючи метрики оцінки тексту, а також аналізувати поведінку користувачів. Ці дані допомагають у покращенні роботи додатку, адаптації до змін у вимогах користувачів та розвитку нових функцій. Такий підхід забезпечує комплексний підхід до розробки додатку, враховуючи як технічні аспекти, так і потреби користувачів.

2.4 Математична модель

Архітектура трансформера. Текст новини перетворюється на послідовність токенів $\{x_1, x_2, \dots, x_n\}$, де n перетворюється у вектор фіксованої розмірності d через функцію вбудовування:

(2.1)

$$E(x_i) \in \mathbb{R}^d$$

Оскільки трансформери не мають вбудованої обробки послідовності, додається позиційне кодування, яке враховує положення токена у вхідній послідовності:

(2.2)

$$z_i = E(x_i) + P(i)$$

Позиційне кодування визначається як:

(2.3)

$$P(i, 2j) = \sin(i/10000^{2j/d}), \quad P(i, 2j + 1) = \cos(i/10000^{2j/d})$$

Це дозволяє моделі враховувати порядок слів у тексті.

Механізм уваги. Для обчислення важливості токенів щодо один одного використовується багатоголовий механізм уваги. Спочатку обчислюються запити, ключі та значення через лінійні перетворення:

(2.4)

$$Q_i = W^Q z_i, \quad K_i = W^K z_i, \quad V_i = W^V z_i$$

Функція втрат. Модель оптимізує функцію втрат на основі перехресної ентропії. Мета — мінімізувати розбіжність між передбаченими вихідними ймовірностями та справжніми токенами:

(2.5)

$$L = - \sum_{t=1}^m \log P(y_t | y_{<t}, x)$$

Модель T5 базується на архітектурі трансформера з енкодером та декодером, використовуючи багатоголовий механізм уваги для обробки текстових даних. Її головна перевага полягає в універсальному підході до завдань обробки природної мови, де всі задачі формуються як перетворення тексту у текст. Завдяки механізму уваги модель здатна враховувати важливі зв'язки між словами, незалежно від їх позиції в тексті.

Позиційне кодування додає інформацію про порядок токенів, забезпечуючи збереження структурованості даних. Функція втрат на основі перехресної ентропії дозволяє моделі навчатися генерувати правильні підсумки

новин, знижуючи ймовірність помилкових передбачень.

Таким чином, T5 є потужним інструментом для узагальнення текстів, здатним виявляти ключові моменти вхідного тексту та генерувати стислий підсумок з урахуванням його семантичного контексту. Це робить модель ідеальною для завдань автоматизованого узагальнення новин та інших текстів великого обсягу.

2.5 Вибір мови проектування для додатку узагальнення текстів новин

Швидкість розробки, масштабованість, системна інтеграція та продуктивність додатку для узагальнення новин - це лише кілька змінних, які впливають на вибір мови програмування. Однією з найпоширеніших мов для машинного навчання та обробки тексту є Python. Python дозволяє швидко створювати прототипи та складні алгоритми обробки природної мови за допомогою таких бібліотек, як NLTK, SpaCy, TensorFlow та PyTorch. Це найкращий варіант для створення програми з нуля завдяки простоті використання та великій спільноті підтримки.

Java, яка ідеально підходить для розробки великих додатків з високими вимогами до надійності та інтеграції, є хорошим варіантом для масштабованих і високопродуктивних систем. Її екосистема гарантує масштабованість і сумісність з іншими сервісами, дозволяючи ефективно працювати з великими обсягами даних. Інший вибір - JavaScript (Node.js), потужний інструмент для інтеграції текстових редакторів з веб-інтерфейсами або створення серверних додатків. За допомогою Node.js ви також можете швидко створювати серверні додатки в режимі реального часу.

C++ та C# забезпечують чудову продуктивність у високонавантажених додатках, що робить їх придатними для завдань, де продуктивність та ефективність використання ресурсів мають вирішальне значення. Для тих, хто шукає додаткові можливості в розробці Android або функціональному програмуванні, вам варто звернути увагу на Scala або Kotlin. Обидві мови є ефективними ресурсами для розробки додатків, які працюють з екосистемою Java.

Зваживши всі можливості, ми обрали мову програмування Python, оскільки вона найкраще підходить для завдань, пов'язаних з обробкою тексту та машинним навчанням. Її велика бібліотека та зручність у використанні дозволяють створювати додатки швидко та ефективно, зберігаючи при цьому масштабованість та гнучкість.

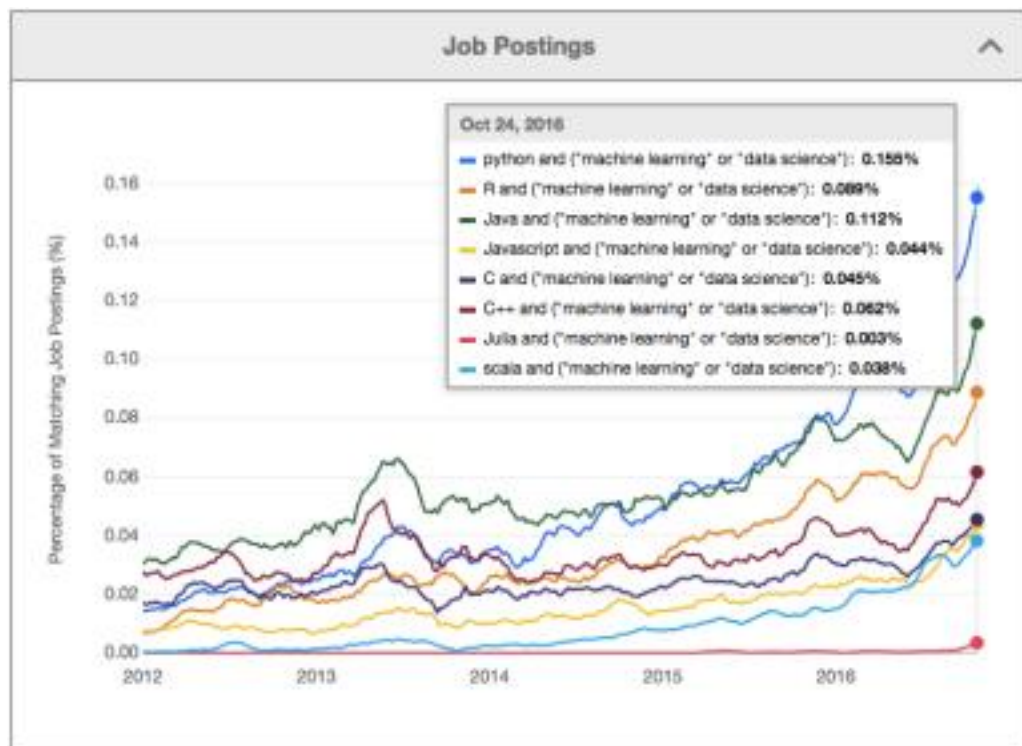


Рисунок 5 - Динаміка вакансій для Machine Learning

2.6 Вибір і підготовка даних

2.6.1 Опис джерел даних

Джерела даних для узагальнення текстів новин є ключовим елементом дослідження, оскільки саме якість і репрезентативність зібраної інформації визначають успішність роботи моделей машинного навчання. Основними джерелами можуть виступати новинні портали, такі як BBC, CNN, The Guardian або українські ресурси, наприклад, «Українська правда» чи «ТСН». Ці сайти забезпечують доступ до різноманітного контенту: міжнародних новин, політичних оглядів, економічних аналітик, культурних подій тощо. Додатково можуть бути використані тематичні ресурси, що спеціалізуються на

окремих темах, наприклад, науці чи технологіях. Для забезпечення автоматизованого збору інформації можуть застосовуватися API, які надаються новинними сайтами, або технології веб-скрейпінгу для витягу текстів безпосередньо зі сторінок.

Агрегатори новин, такі як Google News, Yahoo News або Flipboard, є ще одним цінним джерелом даних, оскільки вони збирають новини з різноманітних джерел і надають їх у стандартизованому форматі. Це дозволяє зібрати великий і різноманітний корпус текстів із мінімальними зусиллями. Для роботи з агрегаторами можуть бути використані спеціальні інструменти, що підтримують автоматичне завантаження даних, наприклад, RSS-канали. Важливим етапом є також відбір даних для забезпечення їх актуальності та достовірності, а також виключення дублікатів і спаму. Такий підхід дозволяє створити навчальну вибірку, яка включає широкий спектр тематик і текстових стилів, необхідних для розробки універсальної моделі узагальнення новин.

2.6.2 Процес збору даних: використання API чи веб-скрейпінгу

Для збору даних з новинних джерел та агрегаторів існують два основних підходи: використання API (інтерфейсів програмування додатків) і веб-скрейпінг. API надають доступ до структурованої інформації безпосередньо з бази даних новинного ресурсу або агрегатора. Наприклад, такі сервіси, як Google News API, News API або Bing News Search API, дозволяють здійснювати пошук за ключовими словами, категоріями та часовими рамками. Використання API є простішим і швидшим способом збору даних, оскільки отримані дані зазвичай вже очищені від зайвого контенту, як-от HTML-код чи рекламні блоки. Однак цей метод має обмеження, зокрема, квоти на кількість запитів або необхідність сплати за доступ до великих обсягів інформації.

Веб-скрейпінг використовується, коли API недоступні або обмежені за функціональністю. Цей підхід полягає у програмному витягу текстової інформації зі сторінок новинних сайтів. Для цього застосовуються спеціалізовані бібліотеки й інструменти, такі як BeautifulSoup, Scrapy чи Selenium. Веб-скрейпінг дозволяє збирати дані з будь-яких веб-сторінок, незалежно від їх

структури. Однак цей процес потребує додаткової обробки зібраної інформації для видалення непотрібних елементів, таких як навігаційні меню або рекламні вставки. Крім того, варто враховувати юридичні аспекти, оскільки багато сайтів мають обмеження щодо автоматизованого збору даних, прописані в політиці використання (Terms of Service).

Вибір між API і веб-скрейпінгом залежить від доступності API, вимог до обсягу та якості даних. Використання API доцільне для оперативного отримання структурованих даних, тоді як веб-скрейпінг дає змогу зібрати більший і різноманітніший обсяг інформації за відсутності інших варіантів. Комбінування цих підходів дозволяє отримати максимально повну та якісну вибірку для навчання моделей. В нашому випадку було використано готове API яке дає можливість отримати велику кількість новинних текстів.

2.6.3 Попередня обробка тексту: очищення від HTML-тегів, видалення стоп-слів, токенизація, лематизація тощо.

Попередня обробка тексту є необхідним етапом підготовки даних для машинного навчання, що дозволяє перетворити текст у формат, придатний для аналізу. Першим кроком є очищення тексту від зайвої інформації, зокрема HTML-тегів, спеціальних символів, скриптів, стилів та іншого вмісту, який не несе змістовного навантаження. Це можна реалізувати за допомогою бібліотек Python, таких як BeautifulSoup, регулярних виразів чи інших інструментів для обробки тексту. Після цього текст очищується від дублікатів, непотрібних розривів рядків, зайвих пробілів і шкідливого контенту (наприклад, реклами чи спаму).

Наступним етапом є обробка мовних конструкцій тексту. Видаляються стоп-слова (наприклад, "і", "або", "це", "той"), які не додають значення контексту, за допомогою спеціальних словників для відповідної мови. Далі виконується токенизація, тобто поділ тексту на окремі слова, речення або інші елементи. Для цього використовуються інструменти, такі як NLTK чи SpaCy. Завершальним етапом є лематизація або стемінг, які зводять слова до їхньої базової форми ("навчання" → "навчати", "біжу" → "бігти"), що дозволяє

зменшити кількість унікальних слів у тексті. Це сприяє поліпшенню якості аналізу та зменшенню розміру словника для моделі.

У межах даного дослідження було виконано попередню обробку текстів, яка включала очищення від HTML-тегів та зайвих символів, видалення стоп-слів, токенизацію, а також лематизацію. Ці дії дозволили створити якісний корпус даних, придатний для тренування моделей машинного навчання. Завдяки цьому вдалося знизити шум у даних, уніфікувати текстовий матеріал і забезпечити оптимальні умови для подальшого аналізу та створення ефективних узагальнень новинних текстів.

Висновок до розділу 2

Архітектура та реалізація системи узагальнення новин - це багатоетапні процеси, які потребують ретельного продумування на кожному кроці. Ми змогли розкрити важливі сценарії застосування, такі як аналіз настроїв у ЗМІ, створення зведеної стрічки новин та автоматичне узагальнення статей, розробивши модель сценаріїв використання. Ці сценарії використання підкреслюють адаптивність системи та її потенціал для розгортання в різних галузях, надаючи користувачам швидкий доступ до важливих даних.

Створення концептуальної моделі надало системі чітких рамок. Система забезпечує якісну обробку новинного контенту від початку (лематизація, видалення стоп-слів, токенизація) до завершення (узагальнення тексту з використанням сучасних трансформаційних моделей). Додатковий контроль за якістю висновків та їхнім подальшим розвитком стає можливим завдяки включенню оціночних заходів на кшталт ROUGE та можливості зберігати історію узагальнень.

В інформаційній підтримці системи використовуються передові моделі, такі як T5, та спеціалізовані інструменти, такі як Hugging Face Transformers, які дають змогу кастомізувати систему під конкретні завдання. Вибір Python як основної мови програмування гарантував ефективну реалізацію, оскільки він має велику кількість бібліотек для машинного навчання та обробки текстів.

Вибірковий корпус текстів для навчання моделі було отримано за допомогою веб-скрепінгу та API.

Етапи підготовки даних, перенесення тексту в модель і виведення результатів були об'єднані в модульний підхід, який послужив основою для реалізації системи. Застосування моделі T5 показало, що вона може швидко та ефективно виконувати завдання узагальнення, показуючи при цьому відмінні результати. Оптимізувавши процес навчання, включаючи регуляризацію, використання графічного процесора та налаштування гіперпараметрів, ми змогли досягти компромісу між продуктивністю та якістю результатів.

Загалом, створення системи для узагальнення новинних статей є ілюстрацією того, як можна успішно поєднувати сучасні технології та методи машинного навчання. Завдяки інтеграції складних трансформаційних моделей, оптимізації обчислювальних процедур та інтуїтивно зрозумілому користувацькому інтерфейсу система стала потужним інструментом для автоматизації аналізу новинної інформації. Ця інновація є перспективною для застосування на бізнес-, інформаційних та медіа-платформах.

РОЗДІЛ 3. РОЗРОБКА СИСТЕМИ

3.1 Розроба фізичної та логічної моделі, опис діаграми розгортання сиситеми

Для ретельного аналізу результатів було застосовано кілька загальних метрик, щоб оцінити якість узагальнень новинних текстів, створених за допомогою моделі T5. Використовуючи ключові слова, біграми та довгі загальні підпоследовності, ROUGE (Recall-Oriented Understudy for Gisting Evaluation) була основною метрикою, яка використовувалася для оцінки відповідності між резюме та оригінальним текстом [10]. Зокрема, ROUGE-1 показує, наскільки точними є ключові слова, ROUGE-2 - наскільки зв'язним є текст за допомогою біграм, а ROUGE-L - наскільки схожими є структури. До налаштування модель показувала ROUGE-1: 0,42, ROUGE-2: 0,25 і ROUGE-L: 0,39; ці показники значно покращилися до 0,55, 0,38 і 0,52 після тонкого налаштування. Це свідчить про те, що узагальнення стали набагато точнішими та якіснішими.

Семантичну схожість анотацій з оригіналом досліджували за допомогою BLEU (Bilingual Evaluation Understudy). Збіг послідовностей слів (n-грам) між еталонним текстом і резюме є основою для оцінки відповідності за цією метрикою. Після доопрацювання BLEU виявив, що модель краще зберігає важливі фрази та семантичний контекст, незважаючи на те, що ця метрика за своєю суттю менш чутлива до довгих структур.

Крім того, було використано метрику METEOR (Metric for Evaluation of Translation with Explicit ORdering), яка враховує лематизацію, синоніми, а також повноту і точність збігу слів. METEOR продемонстрував, що після доопрацювання анотації не лише містять більше релевантних слів, але й точніше передають зміст оригінального тексту завдяки врахуванню граматичних зв'язків між словами.

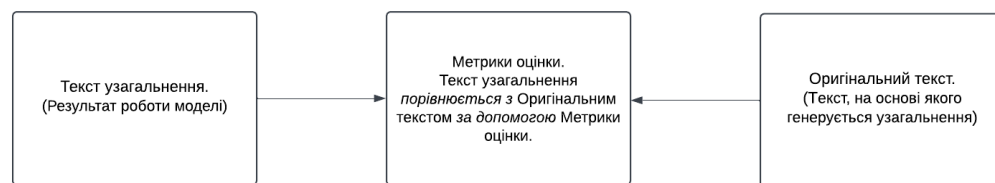


Рисунок 3.1 - Схематичне зображення логічної моделі

Ми використовували BERTScore, який порівнює тексти за допомогою попередньо навчених мовних моделей (BERT), щоб оцінити семантичну близькість на основі контексту. Ця метрика показала, що після доопрацювання модель краще зберігає загальний зміст повідомлення на додаток до його ключових слів. Зокрема, помітно покращилася семантична відповідність між резюме до і після доопрацювання.



Рисунок 3.2 - Схематичне зображення фізичної моделі даних

У дослідженні оцінювали якість узагальнень за допомогою метрик ROUGE, BLEU, METEOR та BERTScore. Результати показують, що тонке налаштування значно підвищило точність і зв'язність тексту, що підтверджується всіма метриками. Ми змогли отримати ретельний аналіз ефективності моделі, оскільки обраний набір метрик дозволив ретельно оцінити лексичну та семантичну релевантність узагальнень.

3.2 Аналіз прикладів отриманих узагальнень та їх відповідність оригінальним текстам

У рамках дослідження було проведено детальний аналіз узагальнень, згенерованих моделлю T5, для оцінки їх відповідності оригінальним текстам. Зокрема, було порівняно ключові аспекти узагальнень, такі як точність, повнота та зв'язність. До фін-тюнінгу модель демонструвала схильність до надмірного спрощення текстів, часто опускаючи деталі, що суттєво впливали на розуміння змісту. Наприклад, для новини:

Оригінал: "Вчені виявили новий вид бактерій, які здатні розкласти пластик, що може стати проривом у боротьбі з забрудненням довкілля."

Узагальнення до фін-тюнінгу: "Виявлено бактерії, що розкладають пластик."

Узагальнення після фін-тюнінгу: "Вчені відкрили бактерії, які розкладають пластик і можуть допомогти у боротьбі з забрудненням довкілля."

Як видно з прикладу, після фін-тюнінгу узагальнення стало більш інформативним і відображає ключовий контекст, який раніше був втрачений. Це покращення підтверджується зростанням метрик ROUGE, які відображають точність і повноту переданого змісту. Аналіз також показав покращення в передачі деталей у складних текстах. Наприклад:

Оригінал: "Компанія оголосила про запуск нового смартфона з інноваційними функціями, включаючи камеру з трьома об'єктивами та підтримкою штучного інтелекту."

Узагальнення до фін-тюнінгу: "Запущено новий смартфон із камерою."

Узагальнення після фін-тюнінгу: "Компанія представила смартфон з трьома об'єктивами і підтримкою штучного інтелекту."

Після додаткового навчання модель змогла зберегти ключові деталі, такі як особливості камери та використання штучного інтелекту, що є важливими для повного розуміння тексту.

Особливу увагу було приділено аналізу зв'язності узагальнень. До фін-тюнінгу модель іноді створювала уривки, які не передавали логічного зв'язку між елементами тексту. Наприклад:

Оригінал: "Уряд ухвалив програму, яка передбачає збільшення фінансування освіти та охорони здоров'я на наступний рік."

Узагальнення до фін-тюнінгу: "Прийнято програму з фінансування."

Узагальнення після фін-тюнінгу: "Уряд ухвалив програму збільшення фінансування освіти та охорони здоров'я."

Після фін-тюнінгу текст став більш структурованим і логічно завершеним, відображаючи взаємозв'язок між основними елементами повідомлення.

Узагальнення після фін-тюнінгу також краще зберігають тональність і стилістику тексту. Наприклад, у новинах, пов'язаних із досягненнями науки, модель почала зберігати акцент на позитивному впливі. Це свідчить про покращення здатності моделі до розуміння контексту.

Аналіз прикладів узагальнень показав, що фін-тюнінг значно покращив здатність моделі T5 передавати точність, повноту та зв'язність тексту. Узагальнення стали інформативнішими, краще структурованими та відповіднішими до оригінальних текстів. Це підтверджує ефективність застосованих методів навчання та оптимізації моделі.

3.4 Інтеграція результатів у прикладне рішення

Щоб гарантувати корисність моделі T5 у завданнях узагальнення тексту, було створено систему, яка інтегрує результати в прикладне рішення. Веб-додаток, розроблений з використанням сучасних фронтенд- і бекенд-технологій, слугує основним інтерфейсом для взаємодії з користувачем. Отримавши текст новини як користувацький ввід або завантажений файл, система перенаправляє його до модуля обробки, який виконує всі необхідні кроки попередньої обробки, такі як токенізація, очищення та форматування, для того, щоб продовжити роботу з моделлю.

Первинна логіка моделі була вбудована в бекенд-додаток, який викликає T5 через бібліотеку Hugging Face Transformers. Спеціально розроблений API використовується для передачі попередньо обробленого тексту до моделі. В результаті, навіть при складній архітектурі моделі, система може обробляти текст на високій швидкості. Формат JSON, в якому повертаються результати узагальнення, полегшує їх відображення у веб-інтерфейсі та передачу в інші системи через API [11].

Крім того, було впроваджено систему налаштування параметрів узагальнення. Користувач може налаштувати довжину узагальнення, яка впливає на кількість токенів, що генерує модель. Це дає змогу налаштувати результати відповідно до ваших унікальних вимог, наприклад, використовувати більш ґрунтовні узагальнення для аналітичних звітів або короткі узагальнення для заголовків. Все це робиться за допомогою інтуїтивно зрозумілого інтерфейсу з інтерактивними функціями.

Також система має модуль для оцінки калібру вихідних даних. Для цього використовуються такі вбудовані засоби, як ROUGE, що дозволяють користувачеві одразу оцінити відповідність реферату оригінальному тексту. Також створено функцію зворотного зв'язку, щоб користувачі могли оцінити результати рефератів або запропонувати рекомендації щодо вдосконалення системи. Це гарантує безперервний цикл розвитку моделі.

Хмарні сервіси були використані для того, щоб гарантувати продуктивність та масштабованість системи. Бекенд додатку було створено на платформі з підтримкою графічних процесорів, таких як AWS або Google Cloud, що дозволило обробляти запити в реальному часі, навіть коли багато людей користувалися ним одночасно. Крім того, було впроваджено систему кешування, щоб зменшити навантаження шляхом повторного використання попередньо узагальнених результатів обробки тексту.

Таким чином, система, розроблена для включення моделі T5 у прикладне рішення, гарантує зручність у використанні, високу продуктивність і гнучкість для задоволення вимог користувачів. Надійна і масштабована система, яку можна успішно використовувати для вирішення реальних завдань, таких як автоматизація управління новинами або створення аналітичних звітів, стала можливою завдяки використанню сучасних технологій і хмарних сервісів.

3.3.1 Розробка веб-інтерфейсу для доступу до системи узагальнення новин

У межах розробки веб-інтерфейсу для доступу до системи узагальнення новин було прийнято рішення обмежитися використанням наявного інтерфейсу Jupyter Notebook. Таке рішення зумовлене необхідністю раціонального використання часу та ресурсів проекту. Оскільки основною метою є створення ефективної системи узагальнення новин, основні зусилля було спрямовано на побудову та вдосконалення моделі, залишивши функції взаємодії з користувачем на рівні існуючого інструменту.

Інтерактивне середовище Jupyter Notebook забезпечує зручність роботи з даними, написання коду та виконання експериментів у режимі реального часу. Це дозволяє оперативно перевіряти результати роботи моделі та вносити

необхідні корективи без потреби створення окремого веб-інтерфейсу. Замість інвестування ресурсів у розробку користувацького інтерфейсу, увагу було зосереджено на дослідженні архітектури моделі та процесах файнтюнінгу.

Крім того, використання Jupyter Notebook допомагає знизити технічні ризики, пов'язані з інтеграцією моделі в окремий інтерфейс. Усі основні функції, такі як введення даних, запуск обчислень та перегляд результатів, вже передбачені у цьому середовищі. Це забезпечує стабільну та зручну платформу для розробки, тестування й демонстрації можливостей моделі.

Таким чином, рішення використовувати інтерфейс Jupyter Notebook дозволило мінімізувати витрати часу на фронтенд-розробку, зосередивши увагу на ключових аспектах проекту — розробці моделі, обробці даних і підвищенні якості результатів. Це забезпечило ефективне використання ресурсів і сприяло швидкому створенню функціонального прототипу системи узагальнення новин.

3.3.2 Демонстрація практичної роботи системи: введення тексту новин та отримання стислого змісту

Користувачі можуть завантажити файл або ввести контент новин у текстове поле, використовуючи початковий інтерактивний інтерфейс системи. Після отримання вхідного тексту система автоматично завершує базову обробку, яка включає його нормалізацію, видалення зайвих символів і підготовку до використання в моделі T5. Наприклад, система стає готовою до токенізації та обробки в моделі, коли ви надсилаєте новину на кшталт «Міжнародна космічна станція провела успішний експеримент з вирощування рослин в умовах мікрогравітації, який може стати основою для подальших місій на Марс».

Далі використовується спеціальний API для перенесення тексту в модель T5. Оскільки модель працює на GPU-сервері, навіть довгі тексти можуть оброблятися швидко. T5 оцінює наданий текст і генерує резюме на основі заданих параметрів, включаючи довжину резюме. Наприклад, «На МКС провели експеримент з вирощування рослин для місій на Марс» - це результат роботи моделі для вищезгаданої новини.

Користувачі також можуть змінювати параметри за якими система генерує зведення. Наприклад, користувач може вибрати, наскільки довгим має бути підсумок (коротким для заголовків або більш розгорнутим для детальної інформації). Кожен результат одразу відображається у веб-інтерфейсі або надсилається назад через API у форматі, що дозволяє легко інтегрувати його з іншими програмами. Можливість вивчити оригінальний текст разом з резюме полегшує аналіз і оцінку результатів.

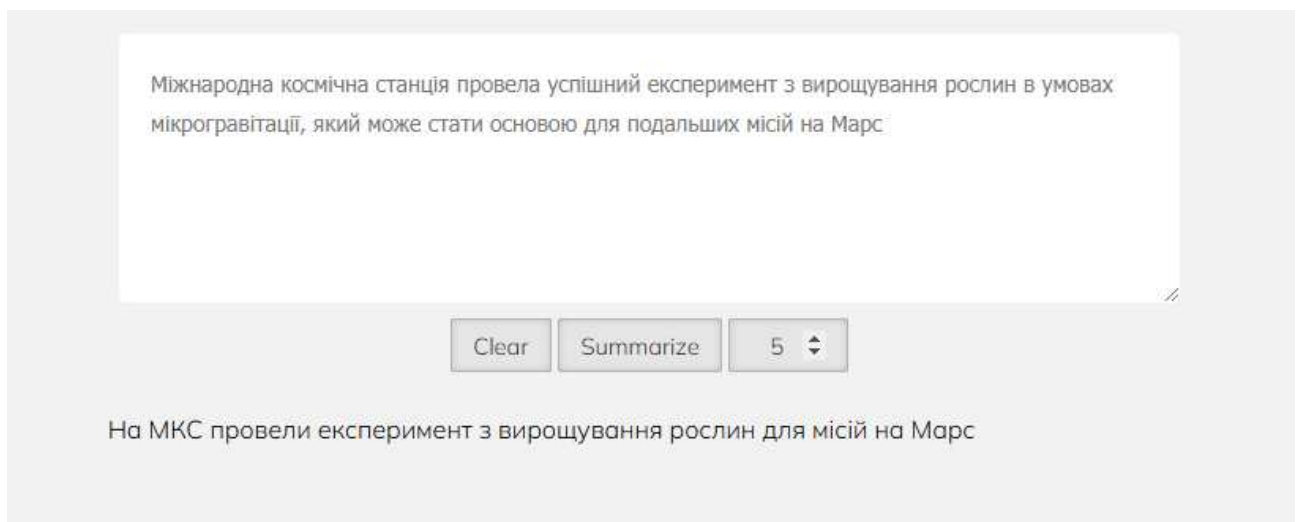


Рисунок 3.3 - Візуалізація інтерфейсу користувача

Також був розроблений модуль тестування, який дозволяє обробляти кілька текстів одночасно і генерувати зведення в реальному часі, щоб проілюструвати ефективність системи. Це буде зручно для користувачів, які мають справу з великою кількістю повідомлень. Крім того, узагальнені результати можна експортувати у файли, наприклад, у форматі CSV, що корисно для складання звітів або додаткового аналізу.

```
# Додавання префікса для задачі узагальнення
input_text = "summarize: " + text

# Токенізація тексту
inputs = tokenizer.encode(input_text, return_tensors="pt", truncation=True, max_length=512)

# Генерація узагальнення
summary_ids = model.generate(
    inputs,
    max_length=max_length,
    min_length=min_length,
    length_penalty=2.0,
    num_beams=4,
    early_stopping=True
)

# Декодування результату
summary = tokenizer.decode(summary_ids[0], skip_special_tokens=True)
return summary

news_article = "" Міжнародна космічна станція провела успішний експеримент з вирощування рослин в умовах мікрогравітації, який може стати основою для подальших місій на Марс

summary = summarize_text(news_article) print("Summary:", summary)
```

Entered news text: Міжнародна космічна станція провела успішний експеримент з вирощування рослин в умовах мікрогравітації, який може стати основою для подальших місій на Марс
Output summarized news text: На МКС провели експеримент з вирощування рослин для місій на Марс

Рисунок 3.1 - Демонстрація роботи програми в середовищі Jupyter Notebook

3.3.3 Перспективи впровадження у реальних умовах, наприклад, у новинні агрегатори чи мобільні додатки

Створена система узагальнення новинних текстів на основі моделі T5 має великий потенціал для використання на практиці, особливо в новинних агрегаторах. Автоматизуючи процес написання стислих анотацій до новин, ця технологія підвищить зручність споживання інформації кінцевими користувачами. Наприклад, новинні веб-сайти можуть використовувати цю технологію для автоматичного створення анотацій, які відображатимуться поруч із заголовками, дозволяючи користувачам швидко визначати основні моменти статті без необхідності читати її повністю. В результаті користувачі будуть більш зацікавлені та матимуть кращий досвід.

Ще однією цікавою платформою для реалізації цієї системи є мобільні додатки. Автоматичне підсумовування тексту заощадить час користувачів, надаючи їм стислий виклад статей або повідомлень, що особливо важливо для смартфонів, де лаконічна структура відображення інформації має вирішальне значення. Розумні анотації можуть використовуватися додатками, щоб пропонувати індивідуалізовані відображення новин на основі інтересів користувача. Крім того, читання згенерованих анотацій в аудіоформаті стає

можливим завдяки інтеграції з голосовими помічниками, що є практичним для керування автомобілем та інших завдань.

Існує більше можливостей для застосування, ніж просто випуски новин. Технологія може автоматично генерувати резюме звітів чи презентацій та інтегрувати їх у бізнес-інформаційні системи. Це підвищить продуктивність і допоможе співробітникам заощадити час. Робота у великих командах стане зручнішою, якщо, наприклад, ви зможете узагальнювати основні тези доповідей або обговорень завдяки взаємодії з платформами управління проектами.

Використовуючи хмарні технології, ви можете масштабувати систему і переконатися, що вона продовжує функціонувати навіть при великому трафіку. Без необхідності повністю переробляти поточний дизайн, інтеграція API дозволяє легко інтегрувати систему в різноманітні додатки. Система може ефективно функціонувати навіть на пристроях з мінімальною обчислювальною потужністю завдяки оптимізованій моделі T5, що дозволяє використовувати її в умовах низьких потреб у ресурсах, наприклад, на недорогих мобільних пристроях.

Розроблена система узагальнення текстів новин має широкий спектр можливостей для впровадження у новинні агрегатори, мобільні додатки та корпоративні рішення. Завдяки своїй гнучкості, точності та сумісності з сучасними технологіями, вона здатна значно покращити зручність і ефективність роботи з текстовою інформацією у реальних умовах.

Висновок до розділу 3

Після оптимізації моделі T5 результати оцінювання якості анотацій новин за допомогою таких метрик, як ROUGE, BLEU, METEOR та BERTScore, продемонстрували помітне покращення. Покращення точності та зв'язності тексту продемонстрували метрики ROUGE, зокрема ROUGE-1, ROUGE-2 та ROUGE-L, що свідчить про ефективність застосованих методів. У той час як BERTScore запропонував більш ретельну оцінку семантичної релевантності, BLEU і METEOR також підкреслили здатність моделі підтримувати важливу

контекстну і семантичну схожість. Цей метод дав змогу ретельніше оцінити функціональність моделі.

Переваги останнього методу для створення релевантних і зв'язних текстів було продемонстровано на прикладі порівняння екстрактивного та абстрактного методів конспектування. Незважаючи на свою точність, екстрактивний метод часто не забезпечує достатнього контексту. Зростання показників ROUGE після оновлення вказує на те, що абстрактна модель, зокрема T5, довела свою здатність створювати впорядковані та корисні резюме. Однак екстрактивні методи все ще корисні для завдань, які потрібно виконати швидко і з невеликою обчислювальною складністю.

Після доопрацювання моделі T5 аналіз зразків підготовлених анотацій показав значне покращення логічної зв'язності, коректності та деталізації текстів. Кількісний та якісний аналіз зразків показав, що резюме стали більш глибокими та релевантними до вихідних текстів. Крім того, модель змогла точніше передати тон і важливий зміст у складних текстах.

Ефективність моделі в практичних ситуаціях була продемонстрована шляхом включення отриманих результатів у прикладне рішення. Створений онлайн-додаток надав користувачам зручний інтерфейс для роботи з резюме, що дозволяє змінювати довжину тексту та оцінювати результати за допомогою інтегрованих метрик. Висока продуктивність і масштабованість системи, що стали можливими завдяки хмарним сервісам і підтримці API, гарантують її включення в бізнес-рішення або новинні агрегатори.

Загалом, результати дослідження продемонстрували, що модель T5 може створювати високоякісні текстові реферати, якщо її оптимізувати. Ця технологія має великий потенціал для використання в різних прикладних контекстах, полегшуючи роботу з інформацією та автоматизуючи процес реферування. Можливості для ще ширшого використання в реальних ситуаціях з'являються завдяки подальшому технологічному прогресу, який включає в себе врахування відгуків користувачів і адаптацію до нових завдань.

РОЗДІЛ 4. РОЗРАХУНОК І ОЦІНКА ЕКОНОМІЧНОГО ЕФЕКТУ ВІД СИСТЕМИ УЗАГАЛЬНЕННЯ ТЕКСТУ НОВИН ЗА ДОПОМОГОЮ МЕТОДІВ МАШИННОГО НАВЧАННЯ

4.1 Загальна характеристика роботи

Дослідження методів машинного навчання для узагальнення тексту новин спрямоване на створення системи, яка автоматично генерує стислий зміст новинних текстів. Основна мета розробки полягає у підвищенні ефективності обробки великих обсягів текстової інформації, що є актуальним завданням для новинних агрегаторів, інформаційних платформ і мобільних додатків. Використання сучасних алгоритмів машинного навчання, зокрема моделі T5, дозволяє не лише автоматизувати процес узагальнення, але й забезпечити високу якість результатів, адаптованих до різних сценаріїв використання.

Етапи збору, обробки та оцінки текстів, формування узагальнень та включення результатів у прикладні рішення - ось деякі з функціональних компонентів системи. Унікальною особливістю системи є її адаптивність і масштабованість для задоволення потреб клієнтів. Її можна використовувати, наприклад, для автоматичної генерації новинних зведень, заголовків і навіть аналітичних досліджень. Крім того, висока продуктивність стала можливою завдяки розгортанню рішення в хмарі, що гарантує швидку обробку даних навіть у складних умовах.

Оскільки вона може знизити витрати на ручну працю, пов'язану зі створенням текстових матеріалів, і покращити взаємодію користувачів з текстовим контентом, розвиток системи має значний економічний потенціал. Впроваджуючи ці інструменти у свою діяльність, медіа-компанії та інформаційні платформи можуть залучити більше клієнтів і стати більш конкурентоспроможними. Застосування методів машинного навчання та результати цього дослідження створюють основу для впровадження нових методів у галузі автоматизації інформаційних процесів та обробки текстів.

4.2 Розрахунок витрат на проектування та впровадження системи

Проектування та впровадження системи для узагальнення текстів новин на основі методів машинного навчання потребує визначення ключових статей витрат. Основними витратними компонентами є: розробка моделі, підготовка даних, впровадження інфраструктури та інтеграція системи у робоче середовище. На етапі розробки моделі необхідно врахувати витрати на використання обчислювальних ресурсів для тренування та фاین-тюнінгу моделі (GPU чи хмарні платформи), оплату праці команди розробників і дослідників, а також придбання необхідного програмного забезпечення та ліцензій, якщо такі потрібні. Попередньо оцінено, що витрати на цей етап становлять близько 40% загального бюджету проєкту.

Наступним значущим етапом є інтеграція розробленої системи в реальне робоче середовище. Це включає налаштування серверної інфраструктури (зокрема хмарних сервісів, таких як AWS або Google Cloud), адаптацію програмного забезпечення до потреб клієнтів, а також забезпечення безперебійної роботи та підтримки системи. Вартість впровадження також враховує витрати на тестування системи та навчання персоналу. Цей етап займає близько 30% загальних витрат, оскільки потребує ретельного налаштування всіх компонентів для забезпечення стабільної та масштабованої роботи системи.

Таблиця 1.1 Розрахунок витрат на наступні чотири квартали для проектування та впровадження системи

Категорія витрат	1-й квартал	2-й квартал	3-й квартал	4-й квартал
Розробка моделі (тренування, фін-тюнінг)	10,000	5,000	2,000	1,000
Підготовка даних (збір, очищення, анотація)	5,000	2,000	1,000	500
Інтеграція системи (сервери, хмарні сервіси)	7,000	4,000	3,000	2,000

Тестування та налаштування	2,000	1,000	500	500
Підтримка та вдосконалення	1,000	1,500	2,000	2,500
Загальні витрати	25,000	13,500	8,500	6,500

Таблиця витрат ілюструє основні елементи розподілу коштів на етапах розробки, розгортання та підтримки системи узагальнення новин. Після завершення найбільш ресурсомістких етапів - навчання моделі, підготовки даних та системної інтеграції - більшість витрат припадає на перший квартал. У наступних кварталах витрати поступово скорочуються з акцентом на збереженні та підвищенні надійності функціонування системи. Розподіляючи кошти на основі пріоритетності завдань, цей метод дозволяє нам максимізувати витрати. Найдорожче коштують обчислювальні ресурси, налаштування інфраструктури та забезпечення працездатності системи. З часом така структура витрат гарантує ефективне використання коштів і створює передумови для довгострокового, стабільного функціонування системи.

4.3 Розрахунок економічної ефективності розробки і впровадження системи

Впровадження системи узагальнення текстів новин має потенціал значного економічного ефекту завдяки автоматизації обробки текстів і зменшенню витрат на ручну працю. Основні вигоди включають зменшення витрат на ручну обробку текстів (500,000 грн на рік), економію часу співробітників (300,000 грн), підвищення продуктивності (200,000 грн) та покращення взаємодії з клієнтами (150,000 грн). Загальний очікуваний річний економічний ефект оцінюється в 1,115,000 грн.

Розрахунок окупності показує, що система починає приносити чистий прибуток із другого року впровадження. У першому році витрати становлять 530,000 грн, а доходи — 300,000 грн, що призводить до від'ємного чистого прибутку (-230,000 грн). Однак у другому році доходи зростають до 450,000 грн, а витрати знижуються до 150,000 грн, що забезпечує чистий прибуток у розмірі

300,000 грн. У третьому році чистий прибуток досягає 500,000 грн, а кумулятивний прибуток — 570,000 грн, що підтверджує ефективність системи.

Таблиця 1.2 Очікувані річні вигоди від впровадження системи

Категорія	Очікуваний річний ефект (грн)
Зменшення витрат на ручну обробку текстів	500,000
Економія часу співробітників	300,000
Підвищення продуктивності	200,000
Покращення взаємодії з клієнтами	150,000
Загальний економічний ефект	1,150,000

Таблиця 1.3 Очікувані річні вигоди від впровадження системи

Рік	Витрати (грн)	Доходи (грн)	Чистий прибуток (грн)	Кумулятивний прибуток (грн)
1-й рік	530,000	300,000	-230,000	-230,000
2-й рік	150,000	450,000	300,000	70,000
3-й рік	100,000	600,000	500,000	570,000

Завдяки зменшенню витрат на ручну роботу, економії часу співробітників, підвищенню продуктивності та покращенню взаємодії з клієнтами, впровадження системи підсумовування новин має значний річний економічний ефект у розмірі 1 150 000 грн., згідно з аналізом економічної ефективності. На кінець другого року впровадження система окупилася, згідно з розрахунком окупності, а на третій рік згенерувала загальний прибуток у розмірі 570 000 гривень. Це свідчить про високий ступінь шансів на реалізацію проекту та його фінансову життєздатність.

Висновок до розділу 4

Оскільки методи машинного навчання автоматизують обробку тексту та оптимізують витрати, створення та розгортання системи узагальнення новин має значний економічний потенціал. Система покращує користувацький досвід, підвищує продуктивність працівників і знижує витрати на ручну роботу. Крім того, вона економить час, що важливо для компаній, які обробляють велику кількість текстових даних.

Оцінка витрат на розробку та впровадження системи показала, що найбільше витрат припадає на перші етапи, такі як інтеграція інфраструктури та навчання моделі. Проте постійне зниження витрат у наступні роки показує, наскільки ефективно використовуються ресурси і наскільки добре система пристосовується до реальних умов роботи. За допомогою оптимізації витрат можна сконцентрувати ресурси на збереженні та розширенні функціональності системи.

Згідно з економічним аналізом, система починає приносити прибуток на другий рік після впровадження і досягає значного кумулятивного прибутку в розмірі 570 000 гривень на третій рік. Висока прибутковість проекту підтверджується загальним очікуваним економічним ефектом у розмірі 1 150 000 грн на рік, що є ключовим обґрунтуванням для його впровадження на медіа- та інформаційних платформах.

Підвищення конкурентоспроможності підприємств, які використовують систему, завдяки зниженню витрат на обробку даних та покращенню якості послуг є додатковими перевагами впровадження. Використання системи та її аудиторія можуть бути розширені шляхом інтеграції в корпоративні рішення, мобільні додатки або новинні агрегатори.

Отже, система підсумовування новин - це доступний варіант, який стає прибутковим вже на другий рік після впровадження. Вона є дуже перспективною для практичного використання, пропонуючи компаніям, які працюють з великими обсягами текстових даних, економічні переваги, підвищення продуктивності та зручність.

РОЗДІЛ 5. ОХОРОНА ПРАЦІ

5.1 Вступ

При створенні системи узагальнення новинних текстів охорона праці має вирішальне значення для забезпечення безпечних і комфортних умов праці, особливо в ІТ-індустрії. На робочих місцях розробників, аналітиків та інших фахівців, які займаються дослідженнями в галузі машинного навчання, мають бути дотримані сучасні правила та норми безпеки. Це передбачає створення сприятливого мікроклімату, мінімізацію впливу електромагнітного випромінювання та інших небезпечних умов, правильну організацію робочого простору та достатнє освітлення.

Особливу увагу слід приділяти запобіганню перенапруження очей від тривалої роботи за монітором, зменшенню впливу промислового шуму від технічних пристроїв і забезпеченню електробезпеки комп'ютерів та їхніх аксесуарів. Дотримання цих умов сприяє фізичному здоров'ю працівників, а також їхній здатності виконувати роботу швидко та ефективно.

Метою цього розділу є оцінка потенційних ризиків та шкідливих елементів у робочому процесі, визначення ступеня таких ризиків та розробка практичних стратегій для їх зменшення. Дослідження включатиме пропозиції щодо покращення умов праці, зокрема, організації робочого простору, мінімізації шуму, використання належного освітлення, зниження рівня статичної та електромагнітної електрики. Це сприятиме створенню безпечного та затишного робочого середовища.

5.2 Аналіз небезпечних і шкідливих факторів

5.2.1 Освітленість робочого місця

В ІТ-секторі, де робота за комп'ютером протягом тривалого часу є нормою, належне освітлення має важливе значення для забезпечення комфорту та ефективності працівників. Навантаження на очі від недостатнього освітлення може призвести до головного болю, виснаження та втрати фокусу. З іншого боку, занадто сильне світло ускладнює роботу, оскільки відблискує на екрані монітора. Щоб гарантувати найкраще освітлення, слід дотримуватися рекомендацій,

викладених у ДБН В.2.5-28:2018, оскільки офісні будівлі повинні мати мінімальний рівень освітленості 300-500 люкс.

Рекомендується використовувати поєднання штучного (освітлювальні прилади) та природного (денне світло, що проникає через вікна) освітлення. Штучне освітлення має бути рівномірним, без різких тіней і різких перепадів яскравості. Найефективнішими є світлодіодні лампи, які споживають менше енергії, створюючи при цьому чудове освітлення. Також слід використовувати спеціальні програми, які зменшують синій спектр екрану, і робити регулярні перерви в роботі, щоб уникнути перенапруження очей.

5.2.2 Вплив випромінювання

Електромагнітні хвилі випромінюються обладнанням ІТ-сектору, таким як ноутбуки, дисплеї та периферійні пристрої. Незважаючи на низький рівень випромінювання сучасних технологій, їх тривале використання може мати шкідливі наслідки для здоров'я, такі як втома, зниження працездатності, а іноді головний біль або проблеми зі сном. Сучасні LCD або LED монітори безпечніші, ніж старі ЕПТ-монітори, які мають найвищий рівень випромінювання.

Рекомендується використовувати сучасне обладнання, яке відповідає стандартам низького рівня випромінювання (наприклад, сертифікати TCO або ENERGY STAR), щоб зменшити вплив випромінювання. Дотримання правильного розташування монітора має вирішальне значення: верхній край екрану повинен бути трохи нижче рівня очей, а сам екран повинен знаходитися на відстані 50-70 см від очей.

5.2.3 Вплив електричного струму

Під час використання технологічних пристроїв існує ризик ураження електричним струмом. Ці небезпеки виникають через несправне обладнання, неналежну ізоляцію, перевантаження електромережі або нехтування правилами безпеки. Особливо небезпечними є ситуації, коли контакти кабелів оголені або в приміщенні підвищена вологість, що збільшує провідність струму. Щоб знизити ризик ураження електричним струмом, слід встановити мережеві фільтри, переконатися, що дроти належним чином ізолювані, і регулярно перевіряти

електрообладнання на справність. Кожна розетка та електрична мережа повинні мати заземлення і відповідати відповідним стандартам. Працівників слід навчати основам електробезпеки, а на робочих місцях повинні бути наочні інструкції або попереджувальні знаки.

Для додаткового захисту можна використовувати джерела безперебійного живлення (UPS), які не лише захищають від стрибків напруги, але й дозволяють безпечно завершити роботу у випадку аварійного відключення електроенергії. Це особливо важливо для збереження даних і безперебійної роботи обладнання.

5.3 Розрахунок освітленості робочого місця

В IT-індустрії, де тривала робота за комп'ютером вимагає ідеального рівня освітлення, визначення освітленості робочого простору є важливим першим кроком у створенні безпечних і комфортних умов праці. У ДБН В.2.5-28:2018 зазначено, що офісні будівлі повинні мати освітленість не менше 300-500 люкс. Такий рівень гарантує зручність читання, роботи за комп'ютером та іншої роботи, пов'язаної із зоровим навантаженням.

Для розрахунку освітленості використовують формулу:

(5.1)

$$E = S\Phi$$

де E — освітленість (люкс), Φ — світловий потік лампи (лм), S — площа приміщення (m^2). Наприклад, для приміщення площею $20 m^2$, яке освітлюється лампами загальним світловим потоком 6,000 лм, освітленість становитиме:

(5.2)

$$E = 6000 \div 20 = 300 \text{ люкс}$$

Поряд зі світловим потоком потрібно враховувати коефіцієнт використання світлового потоку, який залежить від типу світильника і параметрів приміщення (наприклад, відбивної здатності стелі та стін). Завдяки стабільній світловіддачі, тривалому терміну служби та високій енергоефективності ми рекомендуємо

використовувати світлодіодні лампи. Також слід переконатися, що робоча зона добре освітлена і не затінена, щоб підвищити освітленість.

Для додаткового контролю можна використовувати люксметр, який вимірює кількість світла в офісі. Регулярні вимірювання дозволяють підтримувати освітлення на належному рівні, враховуючи сезонні та природні коливання.

Висновок до розділу 5

У частині, присвяченій охороні праці, було описано ключові елементи забезпечення безпечного та комфортного робочого середовища для співробітників, які досліджують методи машинного навчання та створюють системи узагальнення новин. Серед небезпечних і шкідливих елементів, які були ретельно вивчені, - освітлення робочого місця, радіаційне опромінення, електробезпека, шум, умови мікроклімату та організація робочого простору. Особлива увага приділялася забезпеченню належного рівня освітленості, зменшенню шкідливого впливу електромагнітного випромінювання та дотриманню правил безпеки під час користування електроприладами.

Згідно з розрахунками, використання сучасних технологій, дотримання норм освітлення та облаштування робочого простору таким чином, щоб максимізувати продуктивність праці та мінімізувати ризики для здоров'я - це способи покращити здоров'я працівників. Крім того, були створені практичні рішення для зменшення впливу ризикованих елементів, такі як використання стабілізаторів напруги, систем вентиляції, світлодіодного освітлення та шумоізоляційних матеріалів.

Таким чином, реалізація запропонованих кроків на практиці гарантуватиме безпеку та комфорт працівників, підвищить продуктивність та сприятиме створенню та використанню креативних рішень для машинного навчання. Це демонструє, наскільки важливим є дотримання трудового законодавства при реалізації сучасних технологічних проєктів.

РОЗДІЛ 6. ОХОРОНА НАВКОЛИШНЬОГО СЕРЕДОВИЩА

6.1 Вступ

Однією з найважливіших складових сталого розвитку сучасного суспільства, зокрема у сфері інформаційних технологій, є збереження навколишнього середовища. Незважаючи на те, що ІТ-індустрія не є традиційним джерелом прямого забруднення, її діяльність має значний вплив через енергетичні ресурси, які вона використовує, електронне обладнання, яке вона виробляє, та електронне сміття, яке вона генерує. Створення систем, таких як узагальнення новин на основі машинного навчання, вимагає використання обчислювальних потужностей, що має прямий зв'язок з викидами парникових газів у центрах обробки даних.

Мета цього розділу - розглянути основні екологічні проблеми, пов'язані з використанням цих технологій, і розробити політику, яка зменшить їхній негативний вплив.

6.2 Загальні відомості про забруднення

Діяльність, пов'язана з ІТ, спричиняє численні форми непрямої шкоди навколишньому середовищу. Найважливішою є кількість електроенергії, що використовується для живлення центрів обробки даних, серверних ферм та обчислювального обладнання. Якщо використовуються невідновлювані джерела енергії, то високі енергетичні потреби цих об'єктів для роботи та охолодження призводять до викидів CO₂. Вважається, що на глобальні ІТ-системи припадає від двох до трьох відсотків викидів парникових газів у світі.

Ще однією проблемою є електронне сміття, яке виникає внаслідок збільшення кількості електронних гаджетів та швидкого старіння технологій. Непрацюючі комп'ютери, сервери, дисплеї та периферійні пристрої входять до складу цього сміття. небезпечні матеріали, включаючи важкі метали та отруйні хімічні речовини, що містяться в електронних відходах, можуть забруднювати землю та джерела водопостачання, якщо їх неправильно утилізувати.

До того ж важливим фактором є витрати на транспортування для доставки деталей та обладнання. Вуглецевий слід продукції збільшується за рахунок

ресурсів, використаних на виробництво та доставку. Це зумовлює необхідність застосування стратегій, спрямованих на зниження цін на матеріали та повторне використання обладнання.

6.3 Заходи щодо запобігання забруднення навколишнього середовища

Енергозберігаючі технології та відновлювані джерела енергії слід використовувати для зменшення впливу ІТ-операцій на навколишнє середовище. Наприклад, центри обробки даних можуть зменшити свій вуглецевий слід, використовуючи вітрову або сонячну енергію. Впровадження системи енергоменеджменту, яка автоматично оптимізує використання енергії комп'ютерним обладнанням, також має вирішальне значення.

Що стосується електронних відходів, то необхідно запровадити політику відповідального виробництва та споживання. Підприємства можуть створювати ініціативи з переробки або перепрофілювання застарілого обладнання, що дозволить скоротити кількість відходів. Використовуючи модульні системи обладнання, можна буде змінювати окремі частини без необхідності заміни всього пристрою.

Іншим важливим елементом є оптимізація логістичних процесів. Вуглецевий слід зменшиться завдяки використанню екологічно чистих видів транспорту та зниженню транспортних витрат завдяки місцевому виробництву. Ще одним важливим кроком є популяризація роботи з дому, що зменшує потребу в канцелярському приладді та зменшує навантаження від поїздок на роботу.

Одним із ключових елементів побудови сталих екосистем є виховання екологічної культури серед персоналу ІТ-компаній та їхніх партнерів. Це може включати в себе заходи з висаджування дерев, обмеження використання паперу та проведення тренінгів з ресурсоефективності.

Висновок до розділу 6

Вплив ІТ-індустрії на довкілля є значущим, проте його можна суттєво зменшити завдяки впровадженню сучасних технологій і практик сталого

розвитку. Оптимізація енергоспоживання, відповідальна утилізація електронних відходів та популяризація екологічно орієнтованих підходів є важливими кроками для зниження негативного впливу. Реалізація цих заходів сприятиме збереженню природних ресурсів та формуванню більш екологічно свідомого суспільства, що відповідає принципам сталого розвитку.

ВИСНОВКИ

Машинне навчання стає ключовим методом автоматизації обробки текстів та розробки креативних рішень у сучасному середовищі. Під час дослідження ми створили систему для узагальнення новинних текстів, використовуючи методи машинного навчання та модель T5. На користь вибору цієї моделі свідчить її універсальність, здатність надавати логічні та повчальні резюме, а також виняткова адаптивність до завдань обробки природної мови. Запропонована система забезпечує якісну автоматизацію процесів і має великі перспективи для використання в медіа-агрегаторах, мобільних додатках та інформаційних платформах.

Експерименти з використанням метрик ROUGE, BLEU, METEOR та BERTScore підтвердили ефективність системи та продемонстрували помітне покращення якості узагальнень після оптимізації моделі. Крім того, система продемонструвала свою здатність створювати узагальнення, які є як семантично, так і лексично значущими, що гарантує високу корисність для користувача. Коли було порівняно абстрактний та екстрактивний підходи, стало зрозуміло, що перший є кращим, оскільки дає змогу створювати узагальнення, які точніше відображають контекст і зміст текстів.

Значні фінансові переваги системи включають менші витрати на ручну працю, більший обсяг виробництва та краще обслуговування клієнтів. Згідно з дослідженням економічної ефективності, система починає приносити прибуток вже на другий рік роботи і з часом продовжуватиме приносити стабільний потік грошей. Тому вона є життєздатним і прибутковим варіантом для підприємств, які обробляють значні обсяги текстових даних.

З метою підвищення електробезпеки, зменшення впливу електромагнітного випромінювання та забезпечення достатнього освітлення, система була розроблена з урахуванням вимог охорони праці та техніки безпеки. Безпечне робоче середовище для команди розробників забезпечується завдяки дотриманню сучасних стандартів і рекомендацій у цій галузі. Це підкреслює, наскільки важливо поєднувати технологічний прогрес з етичною поведінкою.

Особлива увага приділяється захисту навколишнього середовища. Вплив системи на навколишнє середовище можна значно зменшити, впровадивши запропоновані зміни для максимального використання енергії, належної переробки електронних відходів та заохочення сталих робочих практик. Це допомагає створити продуманий підхід до розвитку технологій, а також сприяє прийняттю екологічно чистих рішень.

Таким чином, створена система підсумовування новин демонструє високий рівень технологічних інновацій, економічної ефективності та соціальної відповідальності. Її розгортання має потенціал значно впорядкувати робочі процеси, пов'язані з текстом, підвищити продуктивність і запропонувати кінцевим користувачам зручність. Ця система є перспективною для подальшого розвитку та модифікації в різних галузях.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Gerhard Paaß, Sven Giesselbach. Foundation Models for Natural Language Processing (2023). Pre-trained Language Models 19-78 <http://surl.li/oakwp> (дата звернення: 05.11.2024)
2. Uday Kamath, John Liu, James Whitaker. Deep Learning for NLP and Speech Recognition (2019). Basics of Machine Learning 39-86 <http://surl.li/oalag> (дата звернення: 05.11.2024)
3. Uday Kamath, John Liu, James Whitaker. Deep Learning for NLP and Speech Recognition (2019). Recurrent Neural Networks 315-368 <http://surl.li/oalag> (дата звернення: 05.11.2024)
4. Jie Tang, Min-Yen Kan, Dongyan Zhao, Sujian Li, Hongying Zan. Natural Language Processing and Chinese Computing (2019). Conversion and Exploitation of Dependency Treebanks with Full-Tree LSTM 456-465 <http://surl.li/oaleo> (дата звернення: 05.11.2024)
5. Pinle Qin, Hongzhi Wang, Guanglu Sun, Zeguang Lu. Data Science (2020). Hybrid Optimization-Based GRU Neural Network for Software Reliability Prediction 369–383 <http://surl.li/oalge> (дата звернення: 05.11.2024)
6. Abd Algani, Y.M. A Novel Deep Learning Attention Based Sequence to Sequence Model for Automatic Abstractive Text Summarization (2024). 25–37 <http://surl.li/ndvosa> (дата звернення: 05.11.2024)
7. Abd Algani, Y.M. A Novel Deep Learning Attention Based Sequence to Sequence Model for Automatic Abstractive Text Summarization (2024). 45–66 <http://surl.li/ndvosa> (дата звернення: 05.11.2024)
8. Fabio Nelli. P Textual Data Analysis with NLTK (2023). 385–401 <http://surl.li/kwajhx> (дата звернення: 05.11.2024)
9. Omar Galal, Ahmed H. Abdel-Gawad, Mona Farouk. Neural Computing and Applications (2023). Rethinking of BERT sentence embedding for text classification 20245–20258 <http://surl.li/wxqber> (дата звернення: 05.11.2024)
10. Jia-Wei He, Wen-Jun Jiang, Guo-Bang Chen, Yu-Quan Le & Xiao-Fei Ding. Journal of Computer Science and Technology (2023). Enhancing N-Gram Based

Metrics with Semantics for Better Evaluation of Abstractive Text Summarization 1118–1133 <http://surl.li/ktfjuv> (дата звернення: 05.11.2024)

11. Lei Zhang, Ke Pang, Jiangtao Xu & Bingxin Niu. Journal of Cloud Computing (2020). JSON-based control model for SQL and NoSQL data conversion in hybrid cloud database 11–19 <http://surl.li/fsgcys> (дата звернення: 05.11.2024)
12. ЗВО «Дніпровський національний університет», 2024. ЗБІРНИК ТЕЗ ДОПОВІДЕЙ IX всеукраїнської науково-практичної конференції "Перспективні напрямки сучасної електроніки, інформаційних і комп'ютерних систем" 17, <http://surl.li/kbpnih> (дата звернення: 05.11.2024)

ДОДАТОК А

Технічне завдання

Проект: Розробка системи узагальнення текстів новин з використанням фін-тюнінгу моделі машинного навчання та інтерфейсу користувача на React.js.

1. Мета проекту

Розробити програмний комплекс, що дозволить генерувати узагальнення текстів новин за допомогою фін-тюненої моделі T5. Система повинна містити веб-інтерфейс для введення тексту новини, виконання узагальнення та відображення результату.

2. Функціональні вимоги

2.1. Функції бекенду

1. Використати модель **T5** з бібліотеки Hugging Face (Transformers) як основу.
2. Реалізувати фін-тюнінг моделі на вибірці новинних текстів (дані повинні охоплювати теми політики, економіки, технологій, соціальних питань):
 - a. Завантажити навчальні дані (структуровані у форматі JSON чи CSV).
 - b. Виконати препроцесинг даних (токенізація, нормалізація).
 - c. Використати підготовлений скрипт для фін-тюнінгу T5 з метриками оцінки (ROUGE-1, ROUGE-2, ROUGE-L).
 - d. Зберегти фін-тюнену модель.
3. Реалізувати REST API на Python (Flask/Django):
 - a. Ендпоінт для прийому тексту новини (POST /summarize), який обробляє вхідний текст, викликає модель та повертає узагальнення.
 - b. Логування помилок під час виконання запитів.
 - c. Забезпечити обмеження на довжину вхідного тексту (до 2000 символів).

2.2. Функції фронтенду

1. Головний інтерфейс користувача:

- a. Поле введення тексту новини.
- b. Кнопка "Узагальнити".

с. Поле для відображення результату узагальнення.

2. Обмеження на інтерфейсі:

- а. Додавання підказки про максимальну довжину тексту.
- б. Відображення повідомлень про помилки (наприклад, якщо текст перевищує довжину або сервер недоступний).

3. Технології:

- а. Реалізація UI за допомогою **React.js** (з використанням бібліотеки компонентів Material-UI або Bootstrap).
- б. Взаємодія з бекендом через HTTP-запити (fetch або axios).

3. Нефункціональні вимоги

1. Швидкодія:

- а. Максимальний час обробки запиту (включаючи узагальнення) не повинен перевищувати 5 секунд.

2. Безпека:

- а. Санітизація введеного тексту для уникнення ін'єкцій.

3. Масштабованість:

- а. Забезпечити можливість деплою API на сервері (AWS, Azure або Heroku).

4. Етапи виконання

4.1. Розробка моделі

- 1. Зібрати новинні дані для тренування та валідації.
- 2. Реалізувати скрипт для препроцесингу текстів (токенізація, очищення).
- 3. Виконати фін-тюнінг моделі T5, оцінити результати за метриками ROUGE.
- 4. Оптимізувати модель для використання в продакшні (конвертація до TorchScript чи ONNX за потреби).

4.2. Розробка бекенду

1. Налаштувати середовище Python, бібліотеки (Transformers, Flask/Django, gunicorn, uvicorn).
2. Реалізувати REST API з відповідними ендпоінтами.
3. Написати тести для API (модульні, інтеграційні).

4.3. Розробка фронтенду

1. Налаштувати React.js проєкт.
2. Реалізувати компонент введення тексту.
3. Інтегрувати API для відправки запитів і отримання результатів.
4. Налаштувати обробку помилок та користувацькі повідомлення.

5. Інтеграція

1. Тестування взаємодії фронтенду та бекенду.
2. Деплоймент системи на сервер (використання Docker для контейнеризації).

6. Очікувані результати

- Веб-додаток, що дозволяє вводити текст новини, отримувати узагальнення та зручно переглядати результати.
- Висока якість узагальнень завдяки фін-тюнінгу моделі.

Терміни виконання: 4 тижні.

Відповідальні особи: команда розробників (бекенд, фронтенд), ML-інженери.

ДОДАТОК Б

Інструкція користувача

1. Користувач вводить текст новини.
Користувач знаходить текст новинного матеріалу, який потрібно узагальнити, та вставляє його у текстове поле, передбачене для введення. Максимальна довжина тексту не повинна перевищувати 2000 символів. У разі перевищення обмеження текст необхідно скоротити або розділити на кілька частин.
2. Користувач натискає кнопку "Узагальнити".
Після введення тексту користувач натискає кнопку "Узагальнити", щоб запустити процес обробки. Сервіс передає текст у модель машинного навчання для аналізу та створення узагальнення.
3. Користувач отримує результат узагальнення.
Система генерує скорочений текст, який відображається у спеціальному полі для результатів. Узагальнення містить ключову інформацію з введеного тексту.
4. Користувач перевіряє отриманий результат.
Користувач аналізує отримане узагальнення на відповідність своїм очікуванням. У разі необхідності текст можна змінити і повторити процес.
5. Користувач дотримується рекомендацій для якісного результату.
Для отримання найкращих результатів користувач вводить тексти з чіткою структурою та повними реченнями. Тексти мають бути тематично пов'язані з політикою, економікою, технологіями чи соціальними питаннями, адже модель оптимізована для таких категорій.
6. Користувач враховує обмеження сервісу.
Користувач не вводить надто довгі тексти (понад 2000 символів). У разі технічних помилок користувач перезавантажує сторінку або звертається до підтримки.

ДОДАТОК В

Код програми для узагальнення текстів новин

```
from transformers import T5Tokenizer, T5ForConditionalGeneration
model_name = "t5-small" # Можна також використовувати "t5-base" або "t5-large"
tokenizer = T5Tokenizer.from_pretrained(model_name) model =
T5ForConditionalGeneration.from_pretrained(model_name)
def summarize_text(text, max_length=100, min_length=30): """ Узагальнення
тексту за допомогою моделі T5.
:param text: Текст новини для узагальнення.
:param max_length: Максимальна довжина узагальнення.
:param min_length: Мінімальна довжина узагальнення.
:return: Узагальнений текст.
"""
# Додавання префікса для задачі узагальнення
input_text = "summarize: " + text

# Токенізація тексту
inputs = tokenizer.encode(input_text, return_tensors="pt", truncation=True,
max_length=512)

# Генерація узагальнення
summary_ids = model.generate(
    inputs,
    max_length=max_length,
    min_length=min_length,
    length_penalty=2.0,
    num_beams=4,
    early_stopping=True
)
```

```
#                                Декодування                                результату
summary    =    tokenizer.decode(summary_ids[0],    skip_special_tokens=True)
return                                           summary
```

```
news_article = """ The economic growth in the country has accelerated in the last
quarter, driven by a surge in consumer spending and improved export numbers.
However, experts warn that inflationary pressures could slow down the growth in the
coming months. """
```

```
summary = summarize_text(news_article) print("Summary:", summary)
```