

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Сумський державний університет
Навчально-науковий інститут бізнесу, економіки та менеджменту СУМДУ
(повна назва інституту/факультету)
Кафедра економічної кібернетики
(повна назва кафедри)

«До захисту допущено»

Завідувач кафедри

_____ В. В. Койбічук

«__» _____ 20__р.

КВАЛІФІКАЦІЙНА РОБОТА
на здобуття освітнього ступеня магістр
(бакалавр / магістр)

зі спеціальності _____ 051 Економіка _____ ,
(код та назва)

Освітньо-професійної програми _____ Економічна кібернетика _____
(освітньо-професійної / освітньо-наукової) (назва програми)

на тему: _____ Розробка модуля для автоматизованого аналізу фінансових звітів компаній з використанням машинного навчання _____

Здобувача (ки) групи ЕК.м-21 _____ Півень Артем Вадимович _____
(шифр групи) (прізвище, ім'я, по батькові)

Кваліфікаційна робота містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело.

(підпис)

_____ Артем Півень _____
(Ім'я та ПРІЗВИЩЕ здобувача)

Керівник

к.е.н., доцентка Койбічук В.В.
(посада, науковий ступінь, вчене звання, Ім'я та ПРІЗВИЩЕ)

(підпис)

Міністерство освіти і науки України
Сумський державний університет
Навчально-науковий інститут бізнесу, економіки та менеджменту
Кафедра економічної кібернетики

ЗАТВЕРДЖУЮ
Завідувач кафедри
к.е.н., доцентка
_____ В.В. Койбічук
“__” _____ 20__ р.

ЗАВДАННЯ НА КВАЛІФІКАЦІЙНУ МАГІСТЕРСЬКУ РОБОТУ
(спеціальність 051 Економіка «Економічна кібернетика»)
студенту 2 курсу, групи ЕК.м- 21

Півень Артем Вадимович

(прізвище, ім'я, по батькові студента)

1. Тема роботи Розробка модуля для автоматизованого аналізу фінансових звітів компаній з використання машинного навчання затверджена наказом по університету від «22» листопада 2023 року № № 1331-VI
2. Термін подання студентом закінченої роботи «__» _____ 20__ року
3. Мета кваліфікаційної роботи полягає у розробці та вдосконаленню методів автоматизованого аналізу фінансових звітів на основі алгоритмів машинного навчання
4. Об'єкт дослідження є процеси автоматизованої обробки та аналізу фінансових звітів з використанням сучасних технологій, базуючись на методах машинного навчання
5. Предмет дослідження є розробка новітніх методів та алгоритмів, які дозволяють підвищити ефективність та точність перетворення фінансового звіту в текстову форму. У цьому контексті особливу увагу приділяється методам глибокого навчання та нейронним мережам, що сприяють автоматизації процесу аналізу фінансових звітів та їх подальшої інтерпретації
6. Кваліфікаційна робота виконується на матеріалах дані, зібрані з різноманітних джерел, включаючи фінансові звіти, наукові публікації, технічні документи, результати експериментів та датасети для тренування та тестування моделей машинного навчання.

7. Орієнтовний план кваліфікаційної роботи, терміни подання розділів керівникові та зміст завдань для виконання поставленої мети

Розділ 1 **ОГЛЯД ПРОБЛЕМИ ГЕНЕРАЦІЇ ТЕКСТОВИХ ДАНИХ – 15.11.2023**
(назва – термін подання)

У розділі 1:

1.1. Загальна характеристика генерації текстових даних на основі аналізу фінансового звіту компаній

1.2. Аналіз сервісів які надають можливість генерації текстових даних на основі аналізу фінансового звіту компаній

(зміст конкретних завдань до розділу, які повинен виконати студент)

Розділ 2 **ОПИС МЕТОДІВ МАШИНОГО НАВЧАННЯ ТА ТЕХНОЛОГІЇ ГЕНЕРАЦІЇ ТЕКСТУ- 25.11.2023**

(назва – термін подання)

У розділі 2:

2.1. Загальна характеристика методів машинного навчання

2.2. Опис відомих технологій генерації тексту за допомогою нейронних мереж

2.3. Можливості машинного навчання в генерації текстових даних на основі аналізу фінансового звіту компаній

(зміст конкретних завдань до розділу, які має виконати студент)

Розділ 3 **РОЗРОБКА МОДУЛЯ ДЛЯ АВТОМАТИЗОВАНОГО АНАЛІЗУ ФІНАНСОВИХ ЗВІТІВ КОМПАНІЙ – 04.12.2023**

(назва – термін подання)

У розділі 3

3.1. Технічне завдання до вирішення поставленої проблем

3.2. Структура і функціонал розробленого модуля автоматизованої системи

3.3. Оцінка очікуваного ефекту від впровадження модуля для автоматизованого аналізу фінансового звіту за допомогою машинного навчання

(зміст конкретних завдань до розділу, які повинен виконати студент)

8. Консультації з роботи:

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
1	Койбічук В.В., к.е.н., доцентка	30.10.2023	06.11.2023
2	Койбічук В.В., к.е.н., доцентка	09.11.2023	23.11.2023
3	Койбічук В.В., к.е.н., доцентка	24.11.2023	08.12.2023

9. Дата видачі завдання: «30» жовтня_2023_року

Керівник кваліфікаційної роботи _____
(підпис)

Койбічук В.В.
(ініціали, прізвище)

Завдання до виконання одержав _____
(підпис)

Півень А.В.
(ініціали, прізвище)

АНОТАЦІЯ

кваліфікаційної магістерської роботи на тему

«РОЗРОБКА МОДУЛЯ ДЛЯ АВТОМАТИЗОВАНОГО АНАЛІЗУ ФІНАНСОВИХ ЗВІТІВ КОМПАНІЙ З ВИКОРИСТАННЯ МАШИННОГО НАВЧАННЯ»

Студента Півень Артема Вадимовича

(прізвище, ім'я, по батькові)

Актуальність теми, обраної для дослідження, визначається тим, що обумовлена декількома ключовими аспектами. По-перше, зростаюча обсяги фінансових даних та їх складність вимагають більш ефективних інструментів для їх обробки та аналізу. По-друге, традиційні методи фінансового аналізу часто вимагають значного часу та ресурсів, що може обмежувати оперативність прийняття рішень. По-третє, машинне навчання надає можливості для автоматизації та оптимізації процесів аналізу, підвищуючи точність та швидкість обробки інформації. Це, у свою чергу, сприяє кращому розумінню фінансового стану компанії та підвищує ефективність управлінського рішення.

Мета кваліфікаційної магістерської роботи полягає у розробці та вдосконаленню методів автоматизованого аналізу фінансових звітів на основі алгоритмів машинного навчання

Об'єктом дослідження є процеси автоматизованої обробки та аналізу фінансових звітів з використанням сучасних технологій, базуючись на методах машинного навчання.

Предметом дослідження є розробка новітніх методів та алгоритмів, які дозволяють підвищити ефективність та точність перетворення фінансового звіту в текстову форму. У цьому контексті особливу увагу приділяється методам глибокого навчання та нейронним мережам, що сприяють автоматизації процесу аналізу фінансових звітів та їх подальшої інтерпретації

Задачами дослідження є

- надати загальну характеристику генерації текстових даних на основі аналізу фінансового звіту компаній;
- проаналізувати сервіси, які надають можливість генерації текстових даних на основі аналізу фінансового звіту компаній;
- надати загальну характеристика методів машинного навчання;
- описати відомі технології генерації тексту за допомогою нейронних мереж;
- розглянути можливості машинного навчання в генерації текстових даних на основі аналізу фінансового звіту компаній;

Для досягнення поставленої мети та задач дослідження були використані такі методи дослідження: У дослідженні задіяні численні методи, зокрема аналіз великих даних, машинне навчання, обробка природної мови, комп'ютерний зір, статистичний аналіз, експериментальний дизайн, крос-валідація та глибоке навчання. Кожен з цих методів відіграє важливу роль у комплексному аналізі та обробці інформації, екстракції знань з фінансових звітів та оптимізації автоматизованих систем генерації тексту.

Інформаційною базою кваліфікаційної магістерської роботи є дані, зібрані з різноманітних джерел, включаючи фінансові звіти, наукові публікації, технічні документи, результати експериментів та датасети для тренування та тестування моделей машинного навчання.

Основний науковий результат кваліфікаційної магістерської роботи полягає у такому: Використання міждисциплінарного підходу у поєднанні теорій та методів з різних наукових областей, що сприяє глибшому розумінню та оптимізації процесів автоматизованої генерації тексту. Це відкриває нові перспективи для подальших досліджень та застосувань у промисловості та академічних колах

Одержані результати можуть бути використані у практичному застосуванні розроблених методів і технологій у реальних проектах. Дослідження демонструє, як

новітні техніки можуть бути інтегровані в існуючі системи обробки даних, значно покращуючи їх продуктивність та точність.

Результати апробації основних положень кваліфікаційної магістерської роботи розглядалися на міжнародній науковій конференції. Назва конференції: «Міжнародна науково-практична конференція «Проблеми та перспективи розвитку фінансово-кредитної системи України 4-5 грудня»».

Назва тези: «Півень А. В., Кушнерьов О. С., Койбічук В. В. Використання машинного навчання для створення текстової інформації на основі аналізу фінансових звітів компаній. Проблеми та перспективи розвитку фінансово-кредитної системи України: Матеріали Міжнародної науково-практичної конференції / за заг. ред.: Л. Л. Гриценко, І. В. Тютюнник. – Суми : Сумський державний університет, 2023. С. 51–54».

Ключові слова: машинне навчання, фінансовий звіт, генерація текстових даних

Зміст кваліфікаційної магістерської роботи викладено на __49__ сторінках.

Список використаних джерел із __73__ найменувань, розміщений на __7__ сторінках.

Робота містить __9__ рисунків, а також __3__ додатків, розміщених на __5__ сторінках.

Рік виконання кваліфікаційної роботи – 2023__ рік.

Рік захисту роботи – 2023 рік

ЗМІСТ:

ВСТУП.....	7
1. ОГЛЯД ПРОБЛЕМИ ГЕНЕРАЦІЇ ТЕКСТОВИХ ДАНИХ	10
1.1 Загальна характеристика генерації текстових даних на основі аналізу фінансового звіту компаній.....	10
1.2 Аналіз сервісів які надають можливість генерації текстових даних на основі аналізу фінансового звіту компаній.....	15
2.. ОПИС МЕТОДІВ МАШИНОГО НАВЧАННЯ ТА ТЕХНОЛОГІЇ ГЕНЕРАЦІЇ ТЕКСТУ	26
2.1 Загальна характеристика методів машинного навчання.....	26
2.2 Опис відомих технології генерації тексту за допомогою нейронних мереж.....	33
2.3 Можливості машинного навчання в генерації текстових даних на основі аналізу фінансового звіту компаній.....	40
3. РОЗРОБКА МОДУЛЯ ДЛЯ АВТОМАТИЗОВАНОГО АНАЛІЗУ ФІНАНСОВИХ ЗВІТІВ КОМПАНІЙ	47
3.1 Технічне завдання до вирішення поставленої проблем	47
3.2 Структура і функціонал розробленого модуля автоматизованої системи	51
3.3 Оцінка очікуваного ефекту від впровадження модуля для автоматизованого аналізу фінансового звіту за допомогою машинного навчання	56
ВИСНОВОК	59
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	60
ДОДАТКИ	67

ВСТУП

Актуальність теми "Розробка модуля для автоматизованого аналізу фінансових звітів компаній з використанням машинного навчання" обумовлена декількома ключовими аспектами. По-перше, зростаюча обсяги фінансових даних та їх складність вимагають більш ефективних інструментів для їх обробки та аналізу. По-друге, традиційні методи фінансового аналізу часто вимагають значного часу та ресурсів, що може обмежувати оперативність прийняття рішень. По-третє, машинне навчання надає можливості для автоматизації та оптимізації процесів аналізу, підвищуючи точність та швидкість обробки інформації. Це, у свою чергу, сприяє кращому розумінню фінансового стану компанії та підвищує ефективність управлінського рішення.

Застосування модулів, розроблених на основі машинного навчання, для аналізу фінансових звітів відкриває широкі можливості для ідентифікації трендів, аномалій та ризиків, що не завжди можуть бути виявлені в процесі традиційного аналізу. Крім того, використання цих технологій дозволяє автоматизувати рутинні процеси аналізу, що звільняє ресурси для зосередження на більш складних завданнях та стратегічному плануванні.

У світлі цих аспектів тема розробки модуля для автоматизованого аналізу фінансових звітів набуває особливої актуальності, оскільки це відповідає поточним вимогам фінансового сектору щодо ефективності, точності та оперативності обробки інформації.

Метою роботи є розробка та вдосконалення методів автоматизованої генерації текстових даних для фінансових звітів на основі алгоритмів машинного навчання.

Завданнями роботи є:

- надати загальну характеристику генерації текстових даних на основі аналізу фінансового звіту компаній;
- проаналізувати сервіси, які надають можливість генерації текстових даних на основі аналізу фінансового звіту компаній;
- надати загальну характеристика методів машинного навчання;
- описати відомі технології генерації тексту за допомогою нейронних мереж;

- розглянути можливості машинного навчання в генерації текстових даних на основі аналізу фінансового звіту компаній;

-

Об'єктом роботи є процеси автоматизованої обробки та аналізу фінансових звітів з використанням сучасних технологій, базуючись на методах машинного навчання.

Предметом роботи є розробка новітніх методів та алгоритмів, які дозволяють підвищити ефективність та точність перетворення фінансового звіту в текстову форму. У цьому контексті особливу увагу приділяється методам глибокого навчання та нейронним мережам, що сприяють автоматизації процесу аналізу фінансових звітів та їх подальшої інтерпретації.

У дослідженні задіяні численні методи, зокрема аналіз великих даних, машинне навчання, обробка природної мови, комп'ютерний зір, статистичний аналіз, експериментальний дизайн, крос-валідація та глибоке навчання. Кожен з цих методів відіграє важливу роль у комплексному аналізі та обробці інформації, екстракції знань з фінансових звітів та оптимізації автоматизованих систем генерації тексту.

Інформаційною базою роботи є дані, зібрані з різноманітних джерел, включаючи фінансові звіти, наукові публікації, технічні документи, результати експериментів та датасети для тренування та тестування моделей машинного навчання.

Наукова новизна отриманих результатів. Використання міждисциплінарного підходу у поєднанні теорій та методів з різних наукових областей, що сприяє глибшому розумінню та оптимізації процесів автоматизованої генерації тексту. Це відкриває нові перспективи для подальших досліджень та застосувань у промисловості та академічних колах. Практичне застосування розроблених методів і технологій у реальних проектах. Дослідження демонструє, як новітні техніки можуть бути інтегровані в існуючі системи обробки даних, значно покращуючи їх продуктивність та точність.

Тема роботи визначила наступну структуру дослідження, що складається із вступу, трьох розділів, висновків, списку використаних джерел.

У першому розділі розкриваються огляд проблеми генерації текстових даних, а саме загальна характеристика генерації текстових даних на основі аналізу фінансового звіту компаній, аналіз сервісів які надають можливість генерації текстових даних на основі аналізу фінансового звіту компаній.

У другому розділі проводиться опис технології для вирішення поставленого завдання, а саме, загальна характеристика методів машинного навчання, опис відомих технологій генерації тексту за допомогою нейронних мереж, можливості машинного навчання в генерації текстових даних на основі аналізу фінансового звіту компаній.

У третьому розділі формується технічне завдання, суть якого полягає в аргументації обраних інструментів та їх аналізі. Також в розділі описана структура модулю та його ключові функції. На завершення в розділі наведено фінальний вигляд модулю та оцінено його роботу.

У висновку відображено основні положення дипломної роботи та сформулювати загальні висновки.

Додаток виносяться розрахункові таблиці та копії фінансової звітності об'єкта дослідження за вибрані для аналізу періоди.

1. ОГЛЯД ПРОБЛЕМИ ГЕНЕРАЦІЇ ТЕКСТОВИХ ДАНИХ

1.1 Загальна характеристика генерації текстових даних на основі аналізу фінансового звіту компаній

Генерація текстових даних є складним процесом, який залучає використання передових технологій у сфері штучного інтелекту. Генерація текстових даних – це процес створення природньомовного тексту, який використовує штучний інтелект (ШІ). Цей процес може включати різні методи та технології, щоб створювати зміст, що здається природнім для людського сприйняття [2].

Вона заснована на методах машинного навчання та глибокого навчання, які аналізують обширні набори текстових даних. Під час аналізу моделі вивчають контекст, граматичні норми, відтінки змісту та інші складності природної мови, завдяки чому можуть створювати зміст, що відображає різні стилістичні, семантичні та контекстуальні особливості.

Генерування тексту відіграє важливу роль у численних сферах, включаючи маркетинг, обслуговування клієнтів, автоматизований переклад та освіту. Воно дозволяє формувати персоналізовані повідомлення, спрощує комунікацію з клієнтами та може забезпечувати навчальний контент відповідно до індивідуальних потреб користувачів. Однак супроводжується це певними викликами, які включають необхідність забезпечення високої точності, адекватності контексту та врахування культурних та етичних аспектів у генерованому контенті [12].

З точки зору етики, генерація тексту накладає відповідальність за використання технології з метою збереження правдивості інформації, запобігання дезінформації та захисту персональних даних. Проблема полягає також у необхідності обмеження зловживань в системах, що можуть генерувати маніпулюючий або шкідливий зміст.

Розробка технологій генерації тексту продовжує еволюціонувати, вимагаючи постійного наукового дослідження для покращення якості, надійності та безпеки сгенерованих текстів. У цьому контексті науковцям необхідно зосереджуватися на

підвищенні ефективності алгоритмів, забезпеченні здатності систем розуміти тонкі нюанси людської мови та врахуванні соціокультурних та етичних аспектів суспільства, у якому ці технології використовуються.

Сучасні системи генерації тексту, як правило, використовують складні моделі машинного навчання, особливо ті, що базуються на глибокому навчанні. Вони навчаються на величезних наборах текстових даних, щоб розуміти контекст, граматику, сленг, емоції та інші аспекти природньої людської мови.

Однією з основних цілей генерації текстових даних є створення тексту, який виглядає достовірно. Текст повинен не тільки слідувати граматичним нормам, але й бути змістовним і релевантним у контексті запиту або заданої теми.

Генерація текстових даних охоплює різні методи та підходи, кожен з яких служить певним цілям та використовується у різноманітних застосуваннях. Ось деякі з основних видів генерації тексту [3]:

1. Синтез мовлення (Text-to-Speech, TTS) – цей метод перетворює текстову інформацію на звукову форму. Він широко використовується в голосових асистентах, системах навігації, а також для забезпечення доступності для осіб з вадами зору.

2. Машинний переклад – технології, які спроможні перекладати текст з однієї мови на іншу автоматично, використовуючи правила граматики, контекст та семантику для забезпечення точності перекладу.

3. Автоматичне створення змісту (Content Generation) – системи, що генерують текст, зазвичай використовуються для автоматичного створення змісту, наприклад, новинних статей, звітів, постів в блогах, на основі набору вхідних даних або шаблонів.

4. Автоматичне реферування – генерація коротких описів або анотацій великих обсягів тексту, як-от наукових статей, книг, документів, для швидкого ознайомлення з основною темою або змістом.

5. Генерація відповідей у діалогових системах – штучний інтелект, який здатний генерувати відповіді у реальному часі в діалогових системах або чат-ботах, щоб спілкуватися або надавати підтримку користувачам.

6. Творча генерація тексту – це включає створення поезії, коротких оповідань, сценаріїв або інших форм художнього тексту за допомогою штучного інтелекту. Цей процес може бути оснований на певному стилі, жанрі або наслідувати відомого автора.

Кожен з цих методів вимагає використання специфічних моделей та алгоритмів машинного навчання, що мають здатність аналізувати, зрозуміти та імітувати людську мову в різних контекстах та для різноманітних цілей.

Генерування тексту може використовуватися в різних сферах, включаючи, але не обмежуючись, чат-ботами, автоматичним створенням змісту для сайтів або публікацій, автоматичним перекладом, синтезом мовлення та навіть у творчих сферах, наприклад, при написанні поезії або прози.

ШІ може генерувати текст, що враховує особисті інтереси, історію браузера або конкретні потреби користувача. Це особливо важливо в маркетингових та рекламних кампаніях, освітніх платформах, а також у взаємодії з персоналізованими віртуальними асистентами [11].

З генерацією текстових даних пов'язані певні етичні питання. Створення вмісту, який здається автентичним, вимагає відповідального підходу до його використання, щоб уникнути дезінформації, зловживань, або маніпулювання думкою громадськості. Також важливо забезпечити конфіденційність і захист даних користувачів.

Незважаючи на значні успіхи, системи генерації тексту все ще мають обмеження у розумінні контексту, сарказму, тонкої емоційної відповіді та інших складних аспектів мовлення. Це ставить певні виклики перед точністю та адекватністю сгенерованого тексту в певних обставинах.

Генерація текстових даних продовжує еволюціонувати з розвитком нових технологій і досліджень у сфері штучного інтелекту, відкриваючи нові можливості та застосування у різних галузях людської діяльності.

Генерація текстових даних на основі аналізу фінансових звітів компаній є процесом, який включає в себе використання технологій штучного інтелекту для обробки, аналізу та інтерпретації великих обсягів фінансової інформації. Цей процес

стає все більш актуальним у контексті зростаючої потреби в автоматизації та оптимізації фінансового аналізу.

Основною метою цього підходу є перетворення складних та часто неструктурованих даних фінансових звітів у зрозумілі, логічно структуровані та легко інтерпретовані текстові висновки. Використовуючи алгоритми машинного навчання та обробки природної мови, системи здатні аналізувати кількісні дані, такі як доходи, витрати, балансові суми, а також якісні показники, включаючи ризики, стратегічні ініціативи та висновки менеджменту.

Важливою складовою цього процесу є здатність системи адаптуватися до специфічних стандартів звітності та фінансових норм, що може значно відрізнятись залежно від країни та галузі діяльності. Це передбачає використання моделей, які враховують міжнародні фінансові стандарти, регуляторні вимоги та індустріальні особливості.

Такі системи не лише підвищують ефективність процесу фінансового аналізу, але й забезпечують більшу прозорість та об'єктивність висновків. Вони можуть використовуватися для автоматичного генерування звітів, що включають синтез фінансових показників, виявлення тенденцій та рекомендацій.

Необхідно зазначити, що хоча ці системи спрощують обробку фінансової інформації, вони потребують регулярного оновлення та налаштування для врахування змін у ринкових умовах, фінансових регуляціях та бізнес-моделях. Таким чином, роль фахівців у фінансовому аналізі залишається важливою, особливо у контексті інтерпретації та використання висновків, отриманих за допомогою штучного інтелекту [16].

Генерація текстових даних на основі аналізу фінансових звітів компаній включає декілька важливих проблем, що виникають внаслідок складності даних, вимог до точності та надійності, а також специфічних викликів, пов'язаних із використанням технологій штучного інтелекту (рис. 1.1).

1. Точність та надійність, це одна з основних проблем, яка полягає у забезпеченні високої точності генерації тексту. Фінансові звіти містять складну та

високоспецифічну інформацію, і невелика помилка у транскрипції або аналізі може призвести до помилкових висновків.

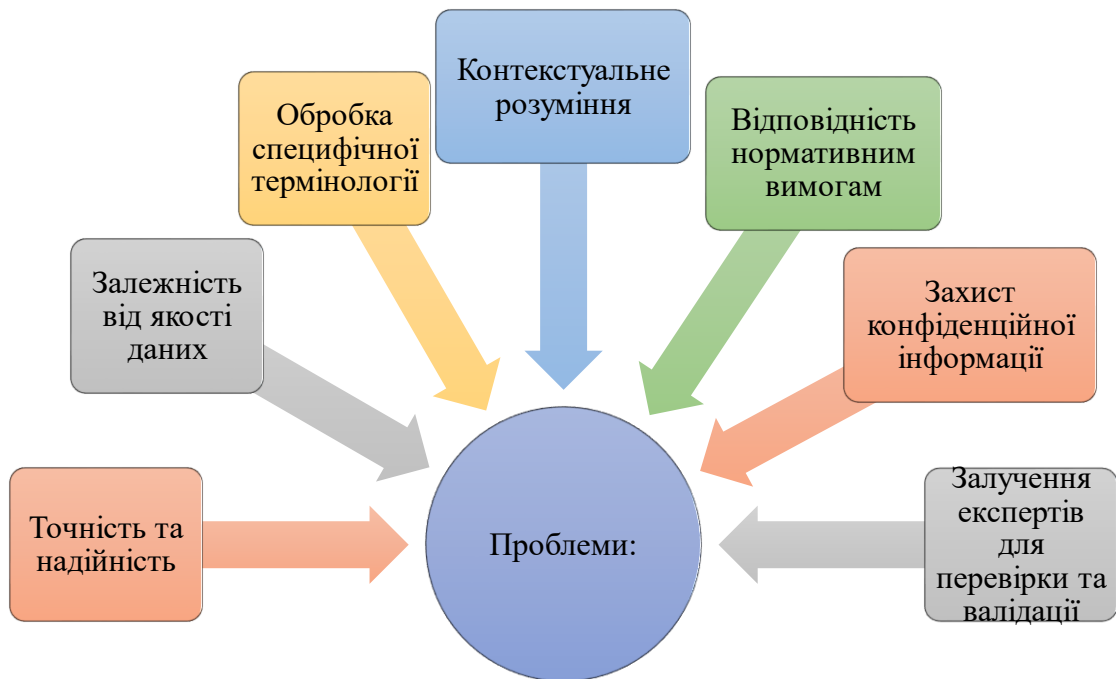


Рисунок 1.1 – Проблеми генерації текстовий даних

2. Залежність від якості даних, бо якість генерованого тексту напряму залежить від якості вхідних даних. Якщо фінансові звіти неповні, містять помилки або представлені у неструктурованому форматі, це може істотно ускладнити процес генерації точного тексту.

3. Фінансові документи часто містять специфічну термінологію, що може бути викликом для систем штучного інтелекту, які не завжди здатні правильно інтерпретувати та використовувати таку спеціалізовану лексику.

4. Автоматизовані системи можуть мати труднощі з контекстуальним розумінням тексту. Вони можуть правильно ідентифікувати ключові цифри та факти, але не завжди здатні правильно інтерпретувати зміст або виявити важливі відносини між різними елементами звіту.

5. З огляду на строгі регуляторні вимоги в сфері фінансової звітності, виникає необхідність забезпечення відповідності автоматично генерованих текстів законодавчим та нормативним стандартам.

6. При роботі з фінансовими даними важливим є забезпечення конфіденційності та безпеки інформації, оскільки вони часто містять чутливі дані.

7. Незважаючи на автоматизацію процесу, потрібно залучати фахівців для перевірки та валідації генерованих текстів, щоб гарантувати їх точність та відповідність стандартам.

Розв'язання цих проблем вимагає постійних досліджень та розробок у галузі обробки природної мови, а також удосконалення алгоритмів машинного навчання, з метою підвищення точності, контекстуальної відповідності та забезпечення надійності генерації текстових даних.

Ці проблеми вимагають уважного балансування між технологічними можливостями генерації тексту та етичними, юридичними та соціальними наслідками їх використання. Розв'язання цих проблем потребуватиме колективних зусиль у сфері регулювання, дослідження та розробки технологій.

1.2 Аналіз сервісів які надають можливість генерації текстових даних на основі аналізу фінансового звіту компаній

Аналіз сервісів, що пропонують генерацію текстових даних на основі аналізу фінансових звітів компаній, виявляє низку ключових аспектів, які важливі для оцінювання їх ефективності та придатності.

Вивчаючи конкретні сервіси, такі як Quill by Narrative Science [17], Yseop [18], або інші платформи, що спеціалізуються на фінансовому аналізі та звітності, важливо оцінювати їх з урахуванням вищезазначених критеріїв. Вибір правильного сервісу може істотно покращити ефективність фінансового аналізу, забезпечуючи цінні інсайти та підтримуючи стратегічне прийняття рішень у компанії.

Станом на квітень 2023 року, існує кілька ключових сервісів та платформ, що спеціалізуються на генерації текстових даних на основі аналізу фінансових звітів компаній. Ці сервіси використовують розширені технології штучного інтелекту для того, щоб автоматизувати та оптимізувати процес аналізу фінансових даних. Деякі з найбільш відомих платформ включають:

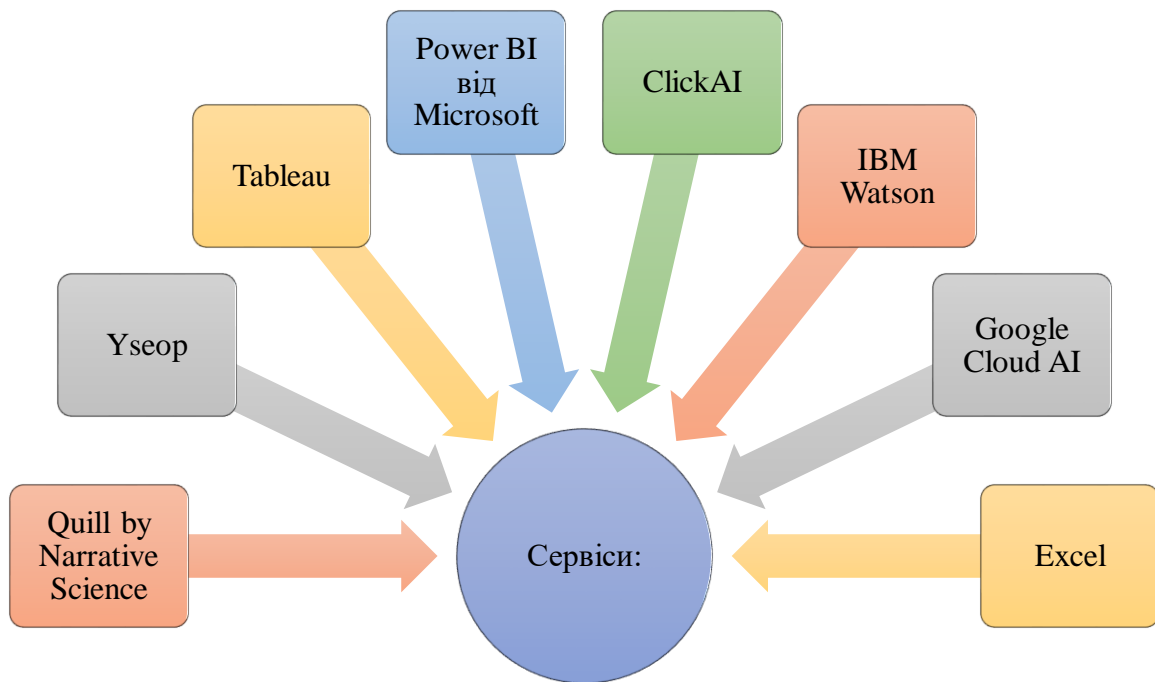


Рисунок 1.2 – Сервіси генерації текстових даних на основі аналізу фінансового звіту компаній

1. Quill by Narrative Science [17]– ця платформа використовує штучний інтелект для автоматичного перетворення даних на зрозумілі текстові звіти. Quill особливо корисний для перетворення складних датасетів на простий для розуміння наратив. Quill by Narrative Science представляє собою платформу штучного інтелекту, яка спеціалізується на автоматичному перетворенні структурованих даних у природномовний текст. Ця технологія заснована на алгоритмах машинного навчання та обробці природної мови і призначена для використання у різних галузях, включаючи фінанси, бізнес-аналітику, маркетинг та інші.

Основна функціональність Quill полягає у здатності аналізувати великі обсяги даних і швидко генерувати з них зрозумілі, інформативні та змістовні текстові звіти. Система спроектована так, щоб імітувати людське розуміння мови, що дозволяє їй створювати тексти, які здаються написаними людиною.

Процес роботи Quill починається з визначення структури та важливих аспектів вхідних даних. Далі система використовує алгоритми для визначення найбільш важливої інформації, яка має бути представлена в тексті. Після цього Quill генерує текст, який може включати описи, висновки, та інші форми текстових представлень даних.

Однією з переваг Quill є його гнучкість і адаптивність. Система може бути налаштована відповідно до специфічних потреб користувача, дозволяючи контролювати стиль, тон і формат генерованих текстів. Це робить Quill зручним рішенням для різних типів бізнес-застосувань, де потрібна швидка і точна інтерпретація даних.

Використання Quill може значно покращити ефективність бізнес-процесів, зокрема у фінансовому аналізі та звітності, забезпечуючи швидке отримання інсайтів з великих наборів даних та звільняючи аналітиків для більш складних завдань. Крім того, автоматизація генерації текстових звітів може знизити витрати та підвищити точність інтерпретації даних.

2. Yseop [18] – цей сервіс автоматизує написання звітів та документації, використовуючи обробку природної мови для генерації детальних звітів з фінансових даних. Yseop ефективний у виявленні ключових інсайтів та тенденцій з великих обсягів даних. Yseop є інноваційним рішенням у сфері штучного інтелекту, яке спеціалізується на автоматизованій генерації тексту та репортингу. Ця платформа використовує технології обробки природної мови (Natural Language Processing, NLP) та машинного навчання для трансформації структурованих даних у детальні, чітко сформульовані текстові звіти.

Основна функція Yseop полягає у тому, що система може автоматично генерувати звіти, аналітичні огляди та документацію на основі великого масиву даних. Це забезпечує значне зниження часу та зусиль, необхідних для ручного аналізу та написання звітів, особливо у сферах, де потрібно швидке та регулярне оновлення інформації, наприклад, у фінансах, фармацевтиці, та інших галузях бізнесу.

Yseop демонструє високу ефективність у розпізнаванні та інтерпретації ключових даних зі складних датасетів. Система здатна не лише виявляти критичні показники та тенденції, але й генерувати розуміння цих даних у вигляді зрозумілого тексту. Така здатність перетворювати числові дані на природномовні висновки робить Yseop цінним інструментом для підтримки прийняття рішень та стратегічного планування.

Yseop відрізняється гнучкістю та масштабованістю, що дозволяє підлаштовувати його під специфічні потреби та вимоги різних організацій. Це включає можливість налаштування стилю, тону та формату генерованого тексту, що робить платформу придатною для широкого спектру застосувань.

Крім того, важливою перевагою Yseop є його відповідність нормам конфіденційності та безпеки даних. Це критично важливо для компаній, які обробляють чутливу інформацію та потребують високого рівня захисту даних.

Загалом, Yseop представляє собою потужний інструмент для автоматизації генерації тексту, що може відігравати ключову роль у підвищенні ефективності бізнес-процесів та оптимізації роботи аналітичних відділів у різних галузях.

3. Tableau [19] більше відомий як інструмент візуалізації даних, він також пропонує можливості з автоматизованого аналізу та інтерпретації даних, що можуть бути використані для створення звітів та дашбордів. Tableau є однією з провідних платформ для візуалізації даних, яка забезпечує користувачам інтуїтивно зрозумілі інструменти для аналізу та інтерпретації великих масивів даних. Хоча основна функція Tableau полягає у візуалізації даних, платформа також може використовуватися для генерації текстових даних на основі аналізу фінансових звітів компаній.

Tableau дозволяє інтегрувати дані з різних джерел, включаючи фінансові звіти, бази даних, та інші системи управління даними. Це дозволяє користувачам агрегувати та об'єднувати фінансові дані для більш глибокого аналізу.

Хоча Tableau не генерує текстові звіти у традиційному розумінні, воно дозволяє користувачам візуалізувати фінансові дані у вигляді графіків, діаграм, карти та інших візуальних форматів. Це може включати тренди доходів, аналіз витрат, порівняння показників ефективності тощо.

Tableau дозволяє створювати інтерактивні звіти та дашборди, які користувачі можуть налаштовувати для виявлення конкретних інсайтів. Ці звіти можуть включати текстові анотації, описи та пояснення, що доповнюють візуальні дані.

Tableau підтримує співпрацю між командами, дозволяючи користувачам ділитися звітами та інсайтами. Це сприяє кращому розумінню фінансових даних і розробці спільних стратегій.

Інтерфейс Tableau розроблено таким чином, щоб зробити аналіз даних доступним навіть для користувачів без глибоких знань у сфері аналітики даних.

У результаті, хоча Tableau може не мати прямої функції генерації текстових даних з фінансових звітів у форматі традиційного звіту, цей інструмент є надзвичайно цінним для візуалізації та глибокого аналізу фінансових даних, що може сприяти кращому розумінню та інтерпретації цих даних.

4. Power BI від Microsoft [20] є потужним інструментом аналізу даних, який надає користувачам можливість не лише візуалізувати, але й аналізувати фінансові дані, а також генерувати з цих даних інформативні звіти. Ця платформа відрізняється гнучкістю та масштабованістю, що дозволяє ефективно використовувати її у різних бізнес-середовищах, від малих підприємств до великих корпорацій.

Power BI надає широкий спектр опцій для візуалізації даних, включаючи графіки, карти, гістограми та інші інструменти, які можна використовувати для створення яскравих та інтуїтивно зрозумілих звітів.

Завдяки інтеграції з різноманітними джерелами даних, Power BI здатний обробляти великі обсяги фінансової інформації, надаючи змогу користувачам глибоко аналізувати фінансові показники, виявляти тенденції та проводити порівняльний аналіз.

Користувачі Power BI можуть створювати інтерактивні дашборди та звіти, які можна легко налаштувати та поділитися з іншими членами команди. Це дозволяє ефективно спілкуватися та приймати рішення на основі актуальних даних.

Power BI включає розширені аналітичні функції, такі як прогнозування, сегментація, кореляційний аналіз та багато іншого, що дозволяє користувачам отримувати більш глибокі інсайти зі своїх даних.

Power BI ефективно інтегрується з іншими продуктами Microsoft, такими як Excel та Azure, а також з широким спектром сторонніх додатків, що робить його гнучким інструментом для різних бізнес-процесів.

Power BI забезпечує високий рівень безпеки даних, включаючи контроль доступу та шифрування, що є критично важливим для обробки фінансової інформації.

Використання Power BI дозволяє компаніям оптимізувати свій фінансовий аналіз, підвищуючи ефективність прийняття рішень та стратегічного планування. Його здатність адаптуватися до різних типів даних та бізнес-вимог робить його цінним інструментом для широкого спектру організацій.

5. ClickAI [21] – цей сервіс забезпечує автоматизацію процесів аналізу даних, включаючи фінансові звіти, з використанням штучного інтелекту. ClickAI представляє собою передовий сервіс, що спеціалізується на автоматизації процесів аналізу даних за допомогою технологій штучного інтелекту. Основною метою цієї платформи є забезпечення користувачам інструментів для ефективного аналізу та інтерпретації великих обсягів даних, включаючи фінансові звіти.

ClickAI використовує алгоритми машинного навчання для автоматизації процесу аналізу даних. Це дозволяє користувачам швидко перетворювати складні датасети у зрозумілі та корисні інформаційні звіти без необхідності глибоких знань у сфері аналітики даних.

Використання ClickAI для аналізу фінансових звітів включає ідентифікацію ключових фінансових показників, аналіз прибутковості, витрат та інших важливих аспектів бізнесу. Це допомагає компаніям отримувати важливі інсайти, що можуть бути використані для підвищення ефективності та прийняття обґрунтованих рішень.

ClickAI забезпечує користувацький інтерфейс, який є зрозумілим та доступним для користувачів без спеціалізованого технічного фону. Це сприяє широкому використанню сервісу в різних бізнес-сферах.

Використання штучного інтелекту для аналізу даних може допомогти в підвищенні точності прогнозів та стратегічного планування, надаючи компаніям змогу краще реагувати на ринкові зміни та оптимізувати свою діяльність.

ClickAI масштабується для обробки великих обсягів даних та може бути адаптований до специфічних потреб та вимог бізнесу.

Загалом, ClickAI відіграє важливу роль у сучасному бізнес-середовищі, дозволяючи компаніям максимально використовувати свої дані для підвищення ефективності, оптимізації процесів та підтримки стратегічного прийняття рішень.

Ці сервіси можуть значно варіюватися за функціональними можливостями, вартістю, легкістю використання та інтеграцією з іншими системами. Вибір конкретного сервісу залежатиме від специфічних потреб бізнесу, обсягу та складності фінансових даних, а також від переваг у візуалізації та аналізі цих даних.

Вибір таких сервісів залежить від багатьох факторів, включаючи точність аналізу, глибину інсайтів, інтеграцію з іншими системами та загальну користувацьку доступність. Ось деякі з ключових характеристик, які слід враховувати:

1. Точність аналізу – сервіси повинні забезпечувати високу точність у виявленні і інтерпретації фінансових даних, у тому числі в розпізнаванні специфічної термінології і складних фінансових концептів.

2. Глибина інсайтів – важливо, щоб платформа здатна не тільки витягувати базову інформацію зі звітів, але й аналізувати тенденції, виявляти ключові показники ефективності (KPI) і надавати глибокі інсайти, які можуть бути корисними для прийняття рішень.

3. Інтеграція з іншими системами. Часто сервіси мають можливість інтегруватися з іншими бізнес-системами, такими як ERP (Enterprise Resource Planning) або CRM (Customer Relationship Management), що дозволяє автоматизувати потоки роботи та покращує загальну ефективність процесів.

4. Інтерфейс користувача має бути інтуїтивно зрозумілим і зручним, щоб користувачі з різним рівнем експертизи могли легко працювати з сервісом.

5. У зв'язку з чутливістю фінансової інформації, сервіси повинні забезпечувати високий рівень безпеки даних та відповідність нормативним вимогам.

6. Скальованість та еластичність – це важливо для компаній, які планують збільшувати обсяги даних або розширювати функціональність з часом.

7. Підтримка та оновлення – надійний сервіс повинен пропонувати ефективну підтримку користувачів та регулярно оновлюватися для відповідності найновішим тенденціям і стандартам.

Станом на квітень 2023 року, на українському ринку можуть бути обмежені можливості спеціалізованих сервісів, які прямо фокусуються на генерації текстових даних на основі аналізу фінансових звітів компаній. Проте, українські компанії та фінансові аналітики можуть використовувати міжнародні платформи та інструменти, адаптовані до своїх потреб. Окрім цього, можливо, існують локальні розробники або стартапи, які пропонують відповідні рішення.

Отже, серед загальновідомих інструментів, які можуть бути доступні та корисні на українському ринку, слід згадати:

1. Tableau та Microsoft Power BI для візуалізації та базового аналізу даних [19].
2. IBM Watson та Google Cloud AI, які мають потужні інструменти для обробки природної мови та можуть бути адаптовані для аналізу фінансових даних [22].

IBM Watson представляє собою передовий набір технологій штучного інтелекту, що включає потужні інструменти для обробки природної мови та машинного навчання. Ці технології можуть бути застосовані для генерації текстових даних на основі аналізу фінансових звітів компаній, надаючи глибокі інсайти та цінну аналітичну інформацію.

IBM Watson може аналізувати великі обсяги фінансових даних, виявляючи ключові показники, тенденції та аномалії. Це включає здатність розпізнавати та інтерпретувати фінансові показники, такі як доходи, витрати, активи, пасиви тощо.

З використанням технологій обробки природної мови, IBM Watson може автоматично генерувати детальні, зрозумілі текстові звіти на основі аналізу фінансових даних. Ці звіти можуть містити описи основних фінансових результатів, порівняльний аналіз з попередніми періодами або конкурентами, а також висновки та рекомендації.

IBM Watson дозволяє користувачам задавати запитання та отримувати відповіді у форматі природної мови, що сприяє кращому розумінню фінансових даних та спрощує процес аналізу.

IBM Watson може інтегруватися з різноманітними джерелами даних, включаючи бази даних, системи ERP (Enterprise Resource Planning), CRM (Customer

Relationship Management) та інші бізнес-системи, що дозволяє забезпечити комплексний аналіз.

Як багатофункціональна платформа, IBM Watson може бути адаптований до специфічних потреб та вимог різних компаній і організацій.

Використання IBM Watson для аналізу фінансових звітів дає можливість компаніям підвищити ефективність своїх аналітичних процесів, отримувати більш глибокі інсайти з даних та сприяє прийняттю обґрунтованих стратегічних рішень на основі точної і актуальної інформації.

Google Cloud AI пропонує широкий спектр інструментів та сервісів, які можуть бути використані для генерації текстових даних на основі аналізу фінансових звітів компаній. Ці інструменти, що базуються на технологіях штучного інтелекту та машинного навчання, дозволяють автоматизувати обробку великих обсягів даних, покращуючи тим самим якість аналітичних звітів та роблячи процес більш ефективним.

Інструменти, такі як Google Cloud Natural Language [23], можуть використовуватися для аналізу текстового вмісту фінансових звітів. Вони можуть виявляти сентимент, виділяти ключові фрази, категоризувати вміст та аналізувати синтаксичну структуру, що допомагає у кращому розумінні контенту звітів.

Інструменти аналітики Google Cloud, такі як BigQuery [24], забезпечують можливість обробки та аналізу великих наборів даних. Це може включати фінансові показники, історичні фінансові дані, ринкові тенденції тощо.

Сервіси, такі як AutoML [25], дозволяють розробникам та аналітикам створювати власні моделі машинного навчання без глибоких знань у цій сфері. Це може бути використано для прогнозування фінансових показників, аналізу ризиків та виявлення аномалій у фінансових даних.

Штучний інтелект Google Cloud може допомогти автоматизувати рутинні завдання, такі як збір та первинна обробка даних, дозволяючи аналітикам зосередитися на більш складних аспектах фінансового аналізу.

Google Cloud AI може бути інтегрований з іншими інструментами та платформами, що використовуються у компанії, що забезпечує єдиний підхід до аналізу даних.

Використовуючи ці інструменти, компанії можуть ефективніше обробляти та аналізувати свої фінансові звіти, отримувати більш глибокі інсайти та приймати обґрунтовані рішення на основі даних. Важливо відзначити, що успішне впровадження цих технологій вимагає відповідних знань та навичок у галузі аналітики даних та штучного інтелекту.

3. Excel [26] з додатковими плагінами для аналізу даних: Багато українських компаній активно використовують Excel для фінансового моделювання, і з допомогою додаткових інструментів ця програма може бути ефективно використана для базового аналізу даних. Microsoft Excel, доповнений спеціалізованими плагінами та додатковими інструментами, є важливим засобом для фінансового аналізу та моделювання в багатьох українських компаніях. Ця програма, широко використовувана через свою універсальність та доступність, може бути значно розширена за допомогою додатків для ефективного аналізу даних. Ось декілька ключових аспектів, які характеризують використання Excel з додатковими інструментами для аналізу фінансових звітів:

Стандартні можливості Excel можуть бути значно розширені за допомогою плагінів та додатків. Це включає інструменти для більш складного аналізу даних, автоматизації процесів, а також розширених візуалізацій.

З допомогою додаткових інструментів, таких як Power Query [27] та Power Pivot [28], Excel може обробляти більші обсяги даних та виконувати складніші аналітичні операції, такі як злиття даних, фільтрація, трансформація та агрегація даних.

Excel дозволяє використовувати різні формули та функції для створення детальних фінансових моделей, включаючи оцінку бюджетів, аналіз рентабельності та прогнозування фінансових показників.

За допомогою Excel можна створювати графіки, діаграми та інші візуальні представлення даних, що полегшує інтерпретацію та аналіз фінансових звітів.

Excel є широко доступним і зрозумілим інструментом для більшості професіоналів, що зменшує поріг входу для виконання складних аналітичних завдань.

Excel дозволяє імпортувати та експортувати дані з різних джерел, що сприяє інтеграції з іншими бізнес-системами та додатками.

Враховуючи ці особливості, Excel з додатковими плагінами та інструментами є важливим ресурсом для українських компаній, що дозволяє їм ефективно аналізувати фінансові дані, виконувати складне фінансове моделювання та приймати обґрунтовані управлінські рішення.

4. Локальні розробки від українських ІТ-компаній. Можливо, існують спеціалізовані рішення, розроблені місцевими компаніями, які пропонують аналіз фінансових звітів.

Важливо відзначити, що українські компанії можуть мати специфічні потреби, пов'язані з місцевим законодавством та особливостями ведення бізнесу, тому при виборі відповідного сервісу слід враховувати ці аспекти. Крім того, можлива потреба в локалізації інтерфейсів та функціональності сервісів для роботи з українською мовою та даними.

2. ОПИС МЕТОДІВ МАШИНОГО НАВЧАННЯ ТА ТЕХНОЛОГІЇ ГЕНЕРАЦІЇ ТЕКСТУ

2.1 Загальна характеристика методів машинного навчання

Машинне навчання (Machine Learning, ML) є підгалуззю штучного інтелекту (AI) [29], яка зосереджується на розробці алгоритмів та статистичних моделей, які дозволяють комп'ютерним системам покращувати свою ефективність автоматично через досвід та використання даних. Машинне навчання використовує різноманітні методи для вирішення специфічних завдань, не вимагаючи явного програмування для кожного конкретного випадку.

Моделі машинного навчання "навчаються" на основі даних. Це означає, що вони аналізують набори даних, виявляють патерни та закономірності, які потім використовуються для прийняття рішень або прогнозів.

Нижче розглянемо типи машинного навчання:

- навчання з учителем (Supervised Learning) [30], ця модель навчається на попередньо розмічених даних, де відомі вхідні та вихідні дані. Ціль полягає у вивченні взаємозв'язків та структур для прогнозування вихідних даних для нових вхідних даних.

- навчання без учителя (Unsupervised Learning) [30], ця модель працює з нерозміченими даними, намагаючись знайти приховані структури або закономірності в даних без попереднього вказівки вихідних результатів.

- частково-наглядове навчання (Semi-Supervised Learning) [31] та навчання з підкріпленням (Reinforcement Learning), це поєднання підходів, де використовуються як розмічені, так і нерозмічені дані, або де модель навчається на основі нагороди за певні дії або рішення.

Методи машинного навчання можуть включати класифікацію (визначення категорії об'єкта), регресію (прогнозування числових значень), кластеризацію

(групування об'єктів за схожими ознаками), зниження розмірності (спрощення даних без втрати важливої інформації), тощо.

Методи машинного навчання застосовуються у різних сферах, від фінансового аналізу до медицини, від маркетингу до рекомендаційних систем.

Машинне навчання включає різні алгоритми, такі як дерева рішень, нейронні мережі, опорні векторні машини, ансамблеві методи тощо, кожен з яких має свої особливості та сфери застосування.

Машинне навчання продовжує розвиватися, відкриваючи нові можливості для автоматизації, підвищення ефективності та отримання глибоких інсайтів з даних у різних областях діяльності.

Отже, методи машинного навчання (ML) становлять фундаментальний компонент сучасних технологій штучного інтелекту (AI) та обробки даних. Вони включають в себе різноманітні алгоритми та технічні підходи, що дозволяють комп'ютерним системам вчитися з даних та вдосконалювати свої функції без явного програмування. Основна характеристика методів машинного навчання включає наступне:

I. Навчання з підкріпленням (Reinforcement Learning) [32]. Навчання з підкріпленням (Reinforcement Learning, RL) являє собою один з основних напрямків у галузі машинного навчання, який зосереджений на розробці алгоритмів, здатних навчатися оптимальним стратегіям поведінки в динамічному середовищі за допомогою системи нагород та покарань. Основна мета навчання з підкріпленням полягає у визначенні послідовності дій, яка максимізує кумулятивну нагороду агента в середовищі.

Алгоритми навчання з підкріпленням зазвичай включають два основних елементи: агента та середовища. Агент виконує дії в середовищі, яке, у свою чергу, надає відповідь на ці дії у вигляді зміни стану середовища та/або надання нагороди агенту. Навчання відбувається через ітераційний процес, де агент намагається оптимізувати свою поведінку для збільшення сумарної нагороди.

Важливим аспектом RL є те, що навчання відбувається без попередньо визначеного набору правил або вказівок. Натомість, агент експериментує з різними

стратегіями та вчиться на основі зворотного зв'язку від середовища, що може включати позитивні або негативні наслідки його дій.

Однією з важливих концепцій в RL є компроміс між дослідженням (exploration) та використанням (exploitation). Дослідження передбачає пробування нових, невикористаних дій з метою виявлення більш ефективних стратегій. Використання означає застосування вже відомих дій, які призводили до отримання нагороди. Ефективний баланс між цими двома стратегіями є ключовим для успішного навчання.

Навчання з підкріпленням знайшло застосування у різноманітних областях, включаючи ігрові системи, автономні транспортні засоби, оптимізацію процесів, робототехніку та інші. Цей підхід відкриває можливості для розвитку систем, здатних самостійно адаптуватися та вдосконалюватися в змінних умовах, що є важливим кроком у розвитку штучного інтелекту.

II. Навчання з учителем (Supervised Learning) [33] представляє собою підхід у сфері машинного навчання, де моделі тренуються на основі набору даних, що містить вхідні приклади та відповідні їм виходи. Основна ідея полягає в тому, що алгоритм намагається вивчити зв'язки між вхідними даними та виходами, щоб згодом здійснювати точні прогнози на нових, нерозмічених даних.

Процедура навчання включає кілька кроків:

1. Підготовка даних: набір даних, який використовується для навчання, складається з прикладів, кожен з яких містить вхідні дані (ознаки) та відповідні виходи (мітки). Важливо, щоб дані були релевантними, достатньо представлені та чистими.

2. Вибір моделі: залежно від характеру задачі та даних вибирається статистична модель або нейронна мережа, яка використовується для навчання.

3. Навчання: процес навчання полягає у використанні алгоритму для аналізу взаємозв'язків між вхідними даними та відповідними виходами. Модель "навчається" на основі цих даних, адаптуючись для забезпечення максимально точного прогнозування.

4. Оцінка: після навчання модель перевіряється на наборі тестових даних, що дозволяє оцінити її точність та надійність у передбаченні.

5. Впровадження: готова модель використовується для роботи з реальними даними, здійснюючи прогнози або визначення необхідних результатів.

Метод навчання з учителем застосовується у двох основних типах задач:

- Класифікація: метою є визначення категорії вхідних даних. Прикладом може бути визначення, чи електронний лист є спамом.

- Регресія: метою є передбачення числових значень на основі вхідних даних. Наприклад, прогнозування цін на нерухомість на основі різних параметрів, таких як розташування та площа.

Таким чином, навчання з учителем відіграє ключову роль у розвитку ефективних алгоритмів машинного навчання, забезпечуючи здатність моделей точно прогнозувати або класифікувати вихідні дані на основі навчальних даних.

III. Навчання без учителя (Unsupervised Learning) [34] являє собою ключовий напрямок у галузі машинного навчання, де моделі тренуються на основі нерозмічених даних. Особливість цього підходу полягає в тому, що моделі намагаються виявити приховані структури, закономірності або взаємозв'язки в даних без явних інструкцій або визначених виходів. Нижче представлено основні аспекти та методи, що характеризують навчання без учителя:

Кластеризація – одна з найбільш поширених задач у навчанні без учителя, де метою є групування об'єктів у кластери на основі схожості їх характеристик. Кластеризація використовується для ідентифікації прихованих патернів у даних, наприклад, у сегментації ринку або організації даних.

Зниження розмірності – метод використовується для зменшення кількості ознак у датасеті, зберігаючи при цьому його основні характеристики. Такий підхід дозволяє спростити моделі та зменшити обчислювальну складність, а також поліпшити інтерпретацію даних. Популярними методами зниження розмірності є головні компоненти аналізу (PCA) та t-розподілене стохастичне вкладення сусідів (t-SNE) [35].

Асоціативні правила – метод використовується для виявлення правил, що описують значущі зв'язки між змінними в наборі даних. Він часто застосовується у

роздрібній торгівлі для аналізу споживацької поведінки та визначення товарів, які часто купуються разом.

Аномальне виявлення – навчання без учителя може бути використане для виявлення аномалій або відхилень в даних. Це має значне застосування в областях, де необхідно виявляти фрод, незвичайні патерни поведінки чи помилки в системах.

Методи навчання без учителя відіграють важливу роль у розумінні та аналізі даних, особливо коли відсутні чіткі вказівки або мітки для тренування моделей. Вони надають можливість ефективно обробляти великі обсяги нерозмічених даних, виявляючи при цьому нові знання та інсайти.

IV. Глибинне навчання (Deep Learning) [36] являє собою важливу підгалузь машинного навчання, яка заснована на використанні архітектур штучних нейронних мереж глибокого навчання. Цей підхід відрізняється від традиційного машинного навчання більшою глибиною та складністю моделей, здатних автоматично виявляти складні закономірності у великих наборах даних.

Основні характеристики глибинного навчання включають:

- глибинне навчання використовує різноманітні типи нейронних мереж, такі як конволюційні нейронні мережі (CNNs) [37] для обробки зображень та відео, рекурентні нейронні мережі (RNNs) [38] для послідовних даних, таких як текст або часові ряди, та інші спеціалізовані архітектури.

- моделі глибинного навчання здатні ефективно обробляти великі набори складних даних, таких як зображення високої роздільної здатності, комплексні звукові доріжки або великі текстові корпуси.

- глибинне навчання автоматично виявляє та використовує характеристики на різних рівнях абстракції, що дозволяє моделям вчиняти більш точні та надійні передбачення чи класифікації.

- глибинне навчання знайшло застосування у багатьох областях, включаючи комп'ютерний зір, обробку природної мови, медичну діагностику, автономні транспортні засоби, розпізнавання мови та багато інших.

– моделі глибинного навчання можуть вчитися з великих датасетів без необхідності детального ручного розмічування, що робить їх особливо ефективними для роботи з неструктурованими або комплексними даними.

– глибинне навчання вимагає значних обчислювальних ресурсів, особливо для тренування великих моделей на великих наборах даних, що часто забезпечується за допомогою спеціалізованого апаратного забезпечення, такого як GPU або TPU [39].

Глибинне навчання продовжує розвиватися та вносить значний вклад у прогрес у сфері штучного інтелекту, відкриваючи нові можливості для розширення можливостей машинного навчання та його застосування у різних галузях.

V. Ансамблеві методи в машинному навчанні [40] представляють собою стратегію, що полягає у комбінуванні декількох навчальних моделей для покращення загальної ефективності, точності та надійності передбачень. Цей підхід ґрунтується на принципі, що ансамбль різноманітних моделей часто працює ефективніше, ніж одиночна модель, навіть якщо ці окремі моделі не є надзвичайно потужними. Два з найвідоміших ансамблевих методів – це Random Forest [41] та Gradient Boosting [42].

1. Random Forest (Випадковий Ліс) [43] використовує множину дерев рішень, кожне з яких навчається на випадково відібраній підмножині тренувального набору даних. Рішення ансамблю визначається шляхом "голосування" або усереднення виходів від окремих дерев. Random Forest ефективний у зменшенні перенавчання (overfitting) та покращенні точності передбачень порівняно з одиночними деревами рішень.

2. Gradient Boosting [44] працює шляхом послідовного додавання нових моделей, які коригують помилки попередніх моделей у ансамблі. Він застосовує градієнтний спуск для оптимізації втрати (loss function) і може бути використаний як для класифікації, так і для регресії. Gradient Boosting зазвичай демонструє високу точність у широкому спектрі застосувань.

Основні переваги ансамблевих методів:

– комбінування декількох моделей часто приводить до кращих результатів, ніж використання однієї моделі.

– ансамблеві методи зменшують ризик перенавчання, особливо у випадку, коли базові моделі схильні до цього.

– ансамблі дозволяють використовувати різні типи моделей та підходів, забезпечуючи більшу гнучкість у вирішенні складних задач.

Ансамблеві методи застосовуються у різних областях машинного навчання, включаючи класифікацію, регресію, виявлення аномалій та інші. Вони є потужним інструментом у арсеналі спеціалістів з даних, дозволяючи побудову більш точних та надійних прогностичних моделей.

VI. Методи машинного навчання відіграють важливу роль у сучасному світі, надаючи значні можливості для автоматизації, оптимізації та підвищення ефективності аналітичних та виробничих процесів. Ці методи використовуються для вирішення широкого спектру задач, від простих до високо складних, і мають ряд ключових переваг:

– машинне навчання може автоматизувати рутинні та часомісткі процеси, такі як введення даних, класифікація об'єктів чи аналіз текстів. Це звільняє час спеціалістів для більш складних та творчих завдань.

– застосування алгоритмів машинного навчання дозволяє компаніям отримувати більш глибокі та точні інсайти з даних, що підвищує якість прийнятих рішень.

– методи машинного навчання здатні обробляти та аналізувати великі набори даних швидше та ефективніше, ніж це можливо вручну.

– використання прогностичних моделей машинного навчання допомагає оптимізувати виробничі лінії, знижуючи витрати та збільшуючи продуктивність.

– машинне навчання забезпечує високу точність у задачах, як-от виявлення дефектів, діагностика захворювань або прогнозування трендів, що є критично важливим у багатьох галузях.

– моделі машинного навчання можуть адаптуватися до нових умов та змін у даних, що дозволяє їм залишатися ефективними у динамічному середовищі.

Завдяки цим перевагам, методи машинного навчання знаходять застосування у різноманітних сферах, включаючи фінанси, охорону здоров'я, виробництво, логістику, маркетинг та багато інших. Вони дозволяють компаніям та організаціям бути більш конкурентоспроможними, інноваційними та ефективними у своїй діяльності.

Методи машинного навчання мають широке застосування у різних сферах, від передбачення поведінки споживачів до медичної діагностики, від фінансового моделювання до автономних транспортних систем. Вони продовжують еволюціонувати, відкриваючи нові можливості для розширення границь того, що можливо з технологіями AI.

2.2 Опис відомих технологій генерації тексту за допомогою нейронних мереж

Генерація тексту за допомогою нейронних мереж стала однією з найбільш вражаючих досягнень у галузі штучного інтелекту та обробки природної мови (NLP). Використання глибинних нейронних мереж дозволяє створювати складні та змістовні текстові послідовності, що знаходять застосування у різних областях, від автоматичного створення контенту до чат-ботів. Ось декілька відомих технологій у цій галузі:

1. Recurrent Neural Networks (RNN) [45] та їх удосконалена версія Long Short-Term Memory (LSTM) [46] мають важливе значення в області обробки послідовних даних, особливо у контексті генерації тексту. Основною особливістю цих мереж є здатність зберігати інформацію про попередній стан, що дозволяє їм ефективно обробляти послідовні дані, такі як текст або часові ряди.

Основною характеристикою Recurrent Neural Networks (RNN) [47] є їх "рекурентна" структура, яка дозволяє передавати інформацію від одного кроку обчислення до наступного. В RNN вихід з попереднього кроку використовуються як вхід для наступного кроку, створюючи внутрішню пам'ять мережі. Це робить RNN

особливо ефективними для обробки послідовних даних, де важливо враховувати контекст та залежності у часі.

Отже, Long Short-Term Memory (LSTM) є варіантом RNN, розробленим для вирішення проблеми "затухання градієнтів" (vanishing gradient problem), яка часто спостерігається в стандартних RNN при обробці довгих послідовностей даних. LSTM-мережі містять спеціальні структури, відомі як "клітини пам'яті" (memory cells), які забезпечують краще збереження довготривалої інформації. Ці клітини регулюють потік інформації за допомогою трьох типів "воріт": вхідних, забуваючих та вихідних, що дозволяє моделі більш ефективно зберігати та використовувати інформацію протягом тривалого часу.

Завдяки цим особливостям, RNN та LSTM знаходять застосування у різноманітних завданнях, пов'язаних з послідовними даними. Вони ефективні у генерації тексту, обробці природної мови, розпізнаванні мовлення, прогнозуванні часових рядів та інших областях, де потрібно розуміння контексту та збереження інформації протягом часу. Їх використання дозволяє створювати більш точні та контекстно залежні моделі для аналізу та генерації послідовних даних.

2. Моделі трансформерів (Transformer Models) [48] відіграють значущу роль у сучасному розвитку області обробки природної мови (NLP). Розроблені дослідниками з Google, трансформери засновані на механізмі уваги (attention mechanism) [49], який дозволяє моделям ефективно обробляти великі блоки тексту, забезпечуючи краще розуміння контексту та зв'язків між словами.

Центральною складовою трансформерів є механізм уваги, який дозволяє моделі зосереджуватися на важливих частинах тексту та ігнорувати менш релевантні деталі. Це покращує здатність моделі до аналізу контексту та зв'язків у тексті.

На відміну від послідовних мереж, таких як RNN та LSTM, трансформери можуть обробляти великі блоки даних паралельно, що значно збільшує швидкість обчислень та ефективність обробки даних.

Трансформери не використовують рекурентну архітектуру, що робить їх менш схильними до проблеми затухання або вибуху градієнтів, які часто спостерігаються в RNN та LSTM.

Трансформери стали основою для розвитку передових моделей у сфері NLP. Найвідоміші серед них – це GPT (Generative Pre-trained Transformer) від OpenAI [50], який ефективно використовується для генерації тексту, та BERT (Bidirectional Encoder Representations from Transformers) від Google [51], який здатен аналізувати зв'язки у тексті у двох напрямках, покращуючи розуміння контексту та семантики.

Завдяки цим характеристикам, трансформери відкрили нові можливості у галузі обробки природної мови, поліпшуючи точність та глибину розуміння текстових даних. Вони застосовуються в широкому спектрі задач, включаючи автоматичний переклад, генерацію тексту, класифікацію текстів, відповіді на запитання та багато інших.

3. GPT-4 (Generative Pre-trained Transformer 4) від OpenAI [52] є наступником моделі GPT-3 [53] та однією з найсучасніших розробок у галузі генерації тексту. Ця модель представляє собою значне вдосконалення попередньої версії, пропонуючи ще більшу потужність, гнучкість та точність у генерації тексту. Основні особливості GPT-4 включають:

GPT-4 має збільшений обсяг та складність порівняно з GPT-3, що дозволяє моделі краще розуміти та генерувати більш складні текстові структури.

Завдяки вдосконаленому механізму обробки природної мови, GPT-4 здатний глибше аналізувати контекст та зв'язки в тексті, що забезпечує більш точні та релевантні відповіді.

Модель знаходить широке застосування в різних сферах, включаючи автоматичне створення контенту, переклад, сумаризацію текстів, відповіді на запитання, а також у сферах, де потрібна творча обробка тексту, наприклад, при створенні літературних творів чи сценаріїв.

GPT-4 здатний взаємодіяти з користувачами в режимі реального часу, адаптуючись до їхніх запитів та потреб, що робить модель особливо корисною у розробці чат-ботів та інтерактивних систем.

Розробники GPT-4 приділили особливу увагу питанням безпеки та етики, намагаючись мінімізувати ризики, пов'язані з неправильним використанням

генеративних моделей, включаючи створення оманливого або неприйняттого контенту.

GPT-4, як і її попередники, є важливою віхою в розвитку технологій обробки природної мови та штучного інтелекту, пропонуючи нові можливості для автоматизації, творчості та інтерактивного спілкування у сфері обробки текстових даних.

4. BERT (Bidirectional Encoder Representations from Transformers) [51] та його варіації внесли значний вклад у галузь обробки природної мови (NLP). Розроблений командою дослідників з Google, BERT використовує трансформерну архітектуру для створення потужних моделей, здатних розуміти контекст та нюанси природної мови. Основні характеристики BERT та його варіацій включають:

На відміну від попередніх моделей, які аналізували текст послідовно (від лівого до правого або навпаки), BERT обробляє весь текст відразу в обох напрямках. Це дозволяє моделі краще розуміти контекст та зв'язки між словами у фразі.

BERT спочатку навчається на великій кількості текстових даних (наприклад, корпусі текстів Вікіпедії), а потім адаптується (тонко налаштовується) до конкретних завдань NLP, таких як класифікація тексту, відповіді на запитання або аналіз емоційного забарвлення.

RoBERTa (Robustly Optimized BERT Pretraining Approach) [54] розроблена Facebook AI, ця модель є варіацією BERT, яка включає зміни в процесі переднього навчання для підвищення ефективності та точності.

DistilBERT [55] модель є спрощеною та оптимізованою версією BERT, яка зберігає більшу частину точності оригінальної моделі, але вимагає значно менше обчислювальних ресурсів. DistilBERT ідеально підходить для ситуацій, де обмежені ресурси або потрібна висока швидкість обробки.

Хоча BERT та його варіації первинно не були розроблені спеціально для генерації тексту, їхні здібності розуміти мову та контекст значно підвищують якість та релевантність генерованого тексту, особливо коли вони використовуються в комбінації з іншими моделями. Ці моделі знаходять застосування в широкому спектрі

задач NLP, від автоматизованої відповіді на запитання до покращення інтерфейсів чат-ботів та систем автоматизованого перекладу.

Ці технології мають різноманітні застосування, від автоматичного генерування відповідей у чат-ботах до створення нових літературних творів. Однак, слід враховувати, що попри високий рівень розвитку, технології генерації тексту все ще мають обмеження, пов'язані з розумінням складних контекстів та збереженням змістовної послідовності.

Окрім вже згаданих методів, таких як RNN, LSTM, трансформери, GPT-3 та BERT, існують інші передові технології та підходи в області машинного навчання та штучного інтелекту, що використовуються для генерації тексту. Деякі з них включають:

1. Архітектура Seq2Seq (Sequence to Sequence) [56] є значущим нововведенням у галузі машинного навчання, яка знайшла широке застосування у сфері обробки природної мови (NLP). Ця модель була спеціально розроблена для перекладу тексту, але її функціональність виявилась ефективною і для інших задач, таких як автоматичне створення тексту. Основні характеристики та застосування архітектури Seq2Seq включають:

Архітектура Seq2Seq складається з двох основних компонентів: кодера (encoder) та декодера (decoder). Кодер обробляє вхідні дані та перетворює їх у внутрішнє представлення, яке потім передається декодеру. Декодер використовує це представлення для генерації вихідного тексту.

У машинному перекладі Seq2Seq моделі навчаються конвертувати текст з однієї мови в іншу. Кодер аналізує вхідний текст на вихідній мові та перетворює його у внутрішнє представлення. Декодер використовує це представлення для створення еквівалентного тексту на цільовій мові.

Seq2Seq моделі також можуть застосовуватися для узагальнення тексту, де вхідний текст конденсується в більш коротку форму, зберігаючи при цьому основний зміст. Цей тип моделей може використовуватися у системах, які генерують відповіді на текстові запитання, перетворюючи запитання на відповідь. Для навчання Seq2Seq

моделей використовуються пари "вхідний текст-вихідний текст". Це дозволяє моделі вчитися відповідним перетворенням, базуючись на навчальних даних.

Seq2Seq моделі є потужним інструментом у NLP, здатним вирішувати складні задачі, пов'язані з перекладом, узагальненням, та генерацією тексту. Вони забезпечують ефективний спосіб обробки послідовних даних, де потрібне розуміння контексту та генерація змістовних відповідей.

2. Conditional Random Fields (CRFs) [57] представляють собою клас статистичних моделей, які знаходять широке застосування в області обробки природної мови (NLP). Ці моделі використовуються для структурованого передбачення, особливо у задачах, де потрібно враховувати контекст та залежності між послідовними даними.

Застосування у розпізнаванні іменованих сутностей (Named Entity Recognition, NER) [58], CRFs ефективно використовуються для виявлення та класифікації іменованих сутностей у тексті, таких як імена осіб, організацій, географічних назв. Вони дозволяють моделі враховувати контекстуальні зв'язки між словами у реченні, що підвищує точність розпізнавання.

Особливістю CRFs є здатність моделювати залежності між послідовними даними. Вони розглядають весь вхідний набір даних (наприклад, речення) як єдине ціле, що дозволяє здійснювати більш точні передбачення на основі загальної інформації.

CRFs навчаються на основі вже розмічених наборів даних, що дозволяє формувати моделі, які вміють визначати патерни та зв'язки в нових текстових даних.

Хоча CRFs самі по собі не є інструментами для генерації тексту, вони можуть використовуватися в комбінації з іншими моделями NLP для покращення якості та точності обробки текстових даних.

CRFs є важливим інструментом в арсеналі фахівців з обробки природної мови, забезпечуючи можливість точного моделювання та аналізу структурованих даних. Їх застосування в NER та інших завданнях, пов'язаних з розумінням тексту, значно підвищує ефективність та точність обробки мовних даних.

3. Autoencoders [59] та Variational Autoencoders (VAEs) [60] є типами нейронних мереж, які використовуються в галузі машинного навчання для зниження розмірності даних та їх генерації. Ці моделі мають унікальні властивості та застосування, особливо у контексті обробки тексту.

Autoencoders – це тип нейронних мереж, який навчається кодувати вхідні дані у компактне представлення та потім реконструювати вихідні дані з цього представлення. Основна мета – зменшення розмірності даних, зберігаючи при цьому їхні ключові характеристики. Autoencoders часто використовуються для виявлення аномалій, шумоподавлення або як частина більш складних систем обробки даних.

Variational Autoencoders (VAEs) – є розширенням звичайних autoencoders, які використовують принципи байєсовської статистики для генерації даних. На відміну від традиційних autoencoders, які лише реконструюють вхідні дані, VAEs здатні генерувати нові дані, які схожі на ті, що використовувалися для навчання. Це робить їх особливо корисними для задач, де потрібно створювати нові зразки даних, наприклад, у генерації тексту.

У контексті обробки тексту, autoencoders та VAEs можуть застосовуватися для створення нових текстів або модифікації існуючих зразків. Наприклад, вони можуть використовуватися для автоматичного переписування текстів з певними змінами у стилі або змісті, генерації нових ідей на основі навчальних текстових даних чи створення нових літературних творів, базуючись на існуючих жанрах.

Завдяки своїй гнучкості та ефективності у виявленні та відтворенні складних шаблонів у даних, autoencoders та VAEs відкривають нові можливості для інновацій у галузі обробки природної мови та створення творчого контенту.

4. Механізм уваги (Attention Mechanism) [61] є інноваційним підходом у сфері машинного навчання, особливо в області обробки природної мови (NLP). Його розробка значно підвищила ефективність нейронних мереж, зокрема у завданнях, пов'язаних з розумінням та генерацією тексту. Основні аспекти та застосування механізму уваги включають:

Механізм уваги дозволяє моделям зосереджуватися на важливих частинах вхідних даних у процесі вироблення виходу. Це особливо корисно в NLP, де необхідно розуміти контекстуальні зв'язки між словами в реченні або абзаці.

Механізм уваги став ключовим компонентом архітектур трансформерів, де він використовується для моделювання залежностей між словами в тексті. Це підвищує здатність моделі до глибшого розуміння мовних структур.

У моделях, що генерують текст, механізм уваги допомагає створювати більш змістовні та контекстно відповідні текстові послідовності, забезпечуючи краще відтворення ідей та збереження зв'язності вмісту.

Хоча механізм уваги став популярним завдяки трансформерам, його можна інтегрувати і в інші типи нейронних мереж, наприклад, у рекурентні нейронні мережі (RNN), для покращення їх здатності до обробки послідовних даних.

Крім завдань у галузі NLP, механізм уваги використовується в багатьох інших застосуваннях, включаючи комп'ютерний зір, прогнозування часових рядів та рекомендаційні системи, де потрібно моделювати складні залежності між даними.

Таким чином, механізм уваги є важливим інструментом у сучасних моделях машинного навчання, забезпечуючи високу гнучкість та ефективність у різних задачах обробки даних.

Ці та інші передові технології відіграють важливу роль у сфері генерації тексту, допомагаючи створювати все більш точні, природні та змістовні текстові послідовності.

2.3 Можливості машинного навчання в генерації текстових даних на основі аналізу фінансового звіту компаній

Машинне навчання (ML) відкриває широкі можливості для генерації текстових даних на основі аналізу фінансових звітів компаній, вносячи значний вклад у процеси автоматизації, підвищення ефективності та якості аналізу. Ось кілька ключових аспектів, як машинне навчання може бути застосоване в цій сфері [62]:

1. Автоматизація первинного аналізу за допомогою алгоритмів машинного навчання стає все більш поширеною практикою в області фінансового аналізу. Цей підхід дозволяє компаніям та фінансовим аналітикам ефективно обробляти великі обсяги фінансових даних, виявляючи ключові показники та витягуючи значущу інформацію. Основні переваги такого підходу включають:

Алгоритми машинного навчання можуть швидко обробляти великі обсяги даних, значно знижуючи час, необхідний для первинного аналізу. Це звільняє час аналітиків для зосередження на більш складних аспектах фінансового аналізу.

Автоматизація допомагає знизити ймовірність помилок, які можуть виникнути при ручному введенні даних. Алгоритми машинного навчання здатні точно ідентифікувати та класифікувати фінансову інформацію.

Машинне навчання може бути використане для ідентифікації та витягування ключових фінансових показників, таких як дохід, витрати, прибутковість, ліквідність та інші важливі метрики.

Автоматизована обробка даних готує фундамент для більш глибокого аналізу, включаючи прогнозування та моделювання фінансових показників.

Машинне навчання дозволяє легко масштабувати процес аналізу, адаптуватися до зростаючих обсягів даних та змін у структурі даних.

Використання машинного навчання для автоматизації первинного аналізу фінансових даних відкриває нові можливості для підвищення продуктивності, точності та глибини фінансового аналізу в компаніях та організаціях.

2. Покращення якості аналізу фінансових звітів за допомогою технік глибинного навчання є одним із ключових напрямків застосування цих технологій у фінансовому секторі. Використання нейронних мереж та інших методів глибинного навчання дозволяє виявити складні закономірності та зв'язки в даних, які можуть бути неочевидними при традиційних підходах аналізу.

Глибинне навчання дозволяє аналізувати великі обсяги фінансових даних, виявляючи складні залежності та взаємозв'язки, які можуть бути прихованими у стандартних фінансових показниках. Це включає в себе аналіз трендів, виявлення аномалій, а також прогнозування фінансових показників.

Нейронні мережі та інші моделі глибокого навчання здатні обробляти не тільки кількісні, але й якісні дані, включаючи текстові звіти та примітки. Це дозволяє отримати глибше розуміння загальної фінансової картини компанії.

Глибинне навчання може використовуватися для прогнозування фінансових показників, що допомагає виявити потенційні ризики та можливості. Це включає в себе прогнозування доходів, витрат, акційних цін та інших критичних фінансових параметрів.

Застосування методів глибокого навчання дозволяє автоматизувати ряд аналітичних задач, таких як класифікація та кластеризація фінансових даних, що сприяє підвищенню ефективності аналітичних відділів.

Шляхом надання більш глибоких та точних аналітичних інсайтів, глибоке навчання підтримує процес прийняття рішень на рівні управління компанією.

Завдяки цим можливостям, глибоке навчання відкриває нові перспективи для фінансових аналітиків та менеджерів, надаючи їм потужні інструменти для більш ефективного та обґрунтованого аналізу фінансових звітів.

3. Генерація звітів та узагальнень за допомогою моделей машинного навчання, особливо тих, що використовують методи обробки природної мови (NLP), є важливим напрямком у сучасному аналізі фінансових даних. Цей підхід може значно спростити та оптимізувати процес підготовки фінансових звітів, забезпечуючи більш високий рівень доступності та зрозумілості інформації для прийняття бізнес-рішень.

Моделі NLP можуть автоматизувати процес створення звітів, витягуючи ключову інформацію з фінансових документів та презентуючи її у структурованому, зрозумілому форматі. Це включає в себе зведення основних фінансових показників, виявлення важливих тенденцій та змін у фінансовому стані компанії.

Моделі машинного навчання можуть використовуватися для узагальнення великих обсягів фінансових даних у короткі висновки. Це допомагає керівництву та інвесторам швидко отримати загальне уявлення про фінансовий здоров'я та ефективність компанії.

Автоматизація генерації звітів забезпечує більш доступну презентацію складної фінансової інформації, що сприяє кращому розумінню та прийняттю рішень на різних рівнях управління.

Моделі машинного навчання можуть інтегруватися з іншими корпоративними системами, такими як системи ERP або CRM, для забезпечення автоматизованого потоку даних та звітності.

Моделі машинного навчання дозволяють кастомізувати формат та зміст звітів залежно від специфічних потреб бізнесу та вимог зацікавлених сторін.

Застосування моделей машинного навчання для генерації звітів та узагальнень у фінансовому аналізі підвищує ефективність обробки інформації, роблячи процес більш автоматизованим, швидким та точним. Це сприяє покращенню якості фінансового планування, аналізу та прийняття рішень на всіх рівнях управління.

4. Прогнозування фінансових тенденцій та розробка прогнозів на майбутнє є ключовими елементами стратегічного планування та управління для будь-якої компанії. Алгоритми машинного навчання, зокрема регресійні моделі, відіграють важливу роль у цьому процесі.

Алгоритми машинного навчання можуть аналізувати історичні фінансові дані, виявляючи тренди та закономірності, які можуть бути неочевидними для людського аналітика. Це включає в себе аналіз циклічних коливань, сезонних змін та інших важливих факторів.

Регресійні моделі та інші аналітичні інструменти машинного навчання можуть використовуватися для прогнозування ключових фінансових показників, таких як доходи, витрати, прибутковість та рентабельність інвестицій.

Моделі машинного навчання допомагають оцінити потенційні ризики та можливості, пов'язані з різними фінансовими стратегіями, шляхом моделювання різних сценаріїв та аналізу їх можливих наслідків [62].

Використання прогностичних моделей дозволяє компаніям приймати більш обґрунтовані рішення щодо інвестицій, визначення бюджетів та інших важливих аспектів фінансового управління.

Машинне навчання може допомогти компаніям швидко адаптуватися до змін у ринкових умовах, аналізуючи поточні тенденції та передбачаючи майбутні зміни.

Застосування машинного навчання для прогнозування фінансових тенденцій та аналізу допомагає компаніям планувати та управляти своєю діяльністю більш ефективно, забезпечуючи конкурентну перевагу та більш стабільне фінансове майбутнє.

5. Індивідуальний підхід до аналізу, який забезпечується за допомогою методів машинного навчання, є значною перевагою для компаній, оскільки дозволяє адаптувати аналітичні рішення до унікальних умов та вимог кожної конкретної організації. Використання машинного навчання для індивідуалізованого аналізу включає наступні аспекти:

Машинне навчання дозволяє розробляти моделі, які враховують унікальні аспекти бізнес-процесів конкретної компанії, такі як циклічність бізнесу, специфіка клієнтської бази або внутрішні процедури управління.

Моделі можуть бути налаштовані таким чином, щоб відповідати структурі фінансової та управлінської звітності компанії, забезпечуючи більш точний та релевантний аналіз.

Машинне навчання дозволяє враховувати специфіку галузі, у якій працює компанія, адаптуючи аналітичні моделі під особливості ринку, конкурентне середовище та регуляторні вимоги.

Моделі машинного навчання можуть бути гнучко адаптовані до змін у бізнесі, таких як розширення діяльності, зміна стратегії або нові ринкові умови, забезпечуючи масштабованість рішень.

Завдяки використанню машинного навчання, компанії можуть оптимізувати свої аналітичні моделі для вирішення конкретних задач, таких як оцінка кредитного ризику, оптимізація ціноутворення або управління запасами [62].

Індивідуальний підхід до аналізу з використанням машинного навчання не тільки підвищує точність та релевантність аналітичних висновків, але й забезпечує компаніям важливу гнучкість, необхідну для адаптації до швидко змінюваного бізнес-середовища.

6. Виявлення аномалій та ризиків за допомогою моделей машинного навчання є критично важливим аспектом у фінансовому аналізі, оскільки це допомагає компаніям вчасно ідентифікувати та реагувати на потенційні проблеми, тим самим запобігаючи фінансовим втратам. Ці моделі використовуються для аналізу великих наборів даних і виявлення відхилень, які можуть вказувати на ненормальну поведінку або ризики.

Моделі машинного навчання можуть виявляти аномальні або нехарактерні транзакції в фінансових звітах, які можуть бути ознаками помилок, шахрайства або інших фінансових проблем.

Застосування машинного навчання дозволяє аналізувати відхилення від звичайних патернів або історичних тенденцій, що може вказувати на потенційні ризики для бізнесу.

Використання прогностичних моделей допомагає фінансовим інститутам оцінювати кредитний ризик та інвестиційні ризики, забезпечуючи краще рішення щодо кредитування та портфельних інвестицій.

Системи на основі машинного навчання можуть постійно моніторити фінансові дані, надсилаючи оповіщення у випадку виявлення аномалій, що дозволяє компаніям швидко реагувати на потенційні проблеми.

Моделі машинного навчання використовуються для ідентифікації ознак шахрайства у фінансових операціях, що допомагає компаніям мінімізувати фінансові збитки від шахрайських дій.

Моделі машинного навчання можуть адаптуватися до змін у фінансовому середовищі та оновлювати свої критерії для виявлення аномалій, забезпечуючи актуальність і точність виявлення ризиків.

Використання машинного навчання для виявлення аномалій та ризиків є потужним інструментом для фінансових аналітиків, який дозволяє підвищити безпеку та надійність фінансових операцій та планування.

Отже, застосування машинного навчання в аналізі фінансових звітів компаній відкриває нові горизонти для більш глибокого та ефективного використання даних,

забезпечуючи компаніям конкурентні переваги та підтримуючи прийняття обґрунтованих рішень.

3. РОЗРОБКА МОДУЛЯ ДЛЯ АВТОМАТИЗОВАНОГО АНАЛІЗУ ФІНАНСОВИХ ЗВІТІВ КОМПАНІЙ

3.1 Технічне завдання до вирішення поставленої проблем

Генерація та аналіз тексту є складним завданням, яке потребує спеціалізованих моделей штучного інтелекту. Розробка власної моделі з нуля може бути трудомістким процесом, що вимагає значних обчислювальних ресурсів та досвіду. Наприклад, для навчання мовних моделей, як GPT-3, мінімальні системні вимоги включають передові GPU з мінімум 16 ГБ, високопродуктивний процесор, принаймні 32 ГБ оперативної пам'яті, і велику кількість вільного місця для зберігання, як правило, на високошвидкісних SSD. Необхідне також швидке підключення до Інтернету та використання спеціалізованих фреймворків, таких як TensorFlow або PyTorch. Час, необхідний для навчання та реалізації моделі, сильно залежить від масштабу, розміру набору даних і складності задачі, але враховуючи ці мінімальні вимоги, процес може займати від декількох тижнів до декількох місяців.

Згідно з статтею на сайті DOU [63], на тренування моделі GPT-3 було витрачено близько 1,5 мільярда доларів США. Це включає витрати на обладнання, електроенергію та оплату праці інженерів та дослідників, які працювали над проектом. Тренування моделі GPT-3 проводилося на кластері комп'ютерів, що складався з 1024 серверів Nvidia Tesla P100. Кожен сервер був оснащений 16 ГБ оперативної пам'яті та 120 ГБ флеш-пам'яті. Загальна потужність кластера становила 1000 терафлопс. Тренування моделі GPT-3 тривало 345 днів. За цей час модель була обрана на даних, що склалися з 500 мільярдів слів.

Сумуючи вище описане на реалізацію власної моделі необхідні значні як фінансові так і обчислювальні ресурси. Тому було вирішено обрати вже готові моделі та використовувати їх з урахуванням специфіки поставленого завдання. Для використання моделей штучного інтелекту через їх API потрібно враховувати

декілька вагомих факторів. Розглянемо одну з найпопулярніших моделей для аналізу та генерації тексту API якої можна використати, GPT від OpenAI [64].

Для використання API потрібно отримати API key, цей ключ дозволить надсилати запити та отримувати відповіді від моделі GPT тарифний план якої був обраний. Для програмної реалізації завдань роботи підходить модель GPT-3.5 Turbo (рис. 3.1).

GPT-3.5 Turbo models are capable and cost-effective.

`gpt-3.5-turbo` is the flagship model of this family, supports a 16K context window and is optimized for dialog.

`gpt-3.5-turbo-instruct` is an Instruct model and only supports a 4K context window.

[Learn about GPT-3.5 Turbo ↗](#)

Model	Input	Output
<code>gpt-3.5-turbo-1106</code>	\$0.0010 / 1K tokens	\$0.0020 / 1K tokens
<code>gpt-3.5-turbo-instruct</code>	\$0.0015 / 1K tokens	\$0.0020 / 1K tokens

Рисунок 3.1 — Тарифні плани використання моделі [65]

В моделі кожне слово вважається окремим токеном, так як і пробіли та знаки пунктуації. Для аналізу тексту за допомогою моделі `gpt-3.5-turbo-106` з 500 токенів на вході і 500 токенів на виході, вартість буде такою:

- вхідні витрати: \$0.0005;
- вихідна вартість: \$0.0010;
- загальна вартість: \$0.0015.

Отже, один аналіз тексту з відповіддю в цьому сценарії коштуватиме \$0,0015. Що при обмеженій кількості запитів не є значною сумою, але якщо кількість запитів та їх складність збільшується, буде збільшуватись як і відповідь так і ціна за цей запит. Враховуючи те що об'єми фінансових звітів можуть буди значними цей варіант використання аналізу штучним інтелектом не підходить для виконання поставлених задач.

Іншим способом використання штучного інтелекту є інтеграція вже готових бібліотек які дозволяють використовувати штучний інтелект.

Бібліотека GPT4All [66], це рішення з відкритим вихідним кодом, спеціально розроблене для запуску великих мовних моделей (LLM) на локальних процесорах і широкому спектрі графічних процесорів. Вона пропонує користувачам можливість тренувати, розгортати та використовувати ці моделі без будь-яких обмежень, пов'язаних з правами власності. Однією з помітних переваг є сумісність з процесорами споживчого класу, а також підтримка пристроїв M1 macOS, що забезпечує ширший доступ. Ця бібліотека дає користувачам можливість тренувати та розгортати власні мовні моделі, що дозволяє їм пристосовувати їх до своїх унікальних потреб. Постійні оновлення, такі як підтримка GGUF для розширених форматів моделей та підтримка локального виводу на різні графічні процесори через Nomic Vulkan, гарантують, що бібліотека залишається актуальною, адаптивною та здатною задовольнити потреби, що змінюються.

Варто зазначити, що GPT4All пропонує моделі розміром від 3 ГБ до 8 ГБ. Це свідчить про їхні потенційні можливості, але також підкреслює значні ресурси пам'яті, які вони вимагають. Крім того, значні розміри моделей вказують на те, що використання GPT4All може призвести до виконання завдань з інтенсивними обчисленнями, що потребують потужного апаратного забезпечення для забезпечення ефективної роботи. Враховуючи специфіку завдання аналізу фінансової звітності, хоча швидкість тут не є основним критерієм проте можна досягти кращих результатів використовуючи аналогові бібліотеки.

Bard API [67], пакет Python, який отримує відповіді від Google Bard, використовуючи значення файлів cookie. Для встановлення використовується менеджер пакетів Python pip. Для автентифікації метод передбачає вилучення значення cookie-файлу `__Secure-1PSID` з сеансу браузера на офіційному сайті Bard. Це значення не є офіційним ключем API і може часто змінюватися, тому, доведеться регулярно перевіряти і оновлювати його, щоб уникнути помилок.

З точки зору використання, Bard API надає простий спосіб надсилати текст до Bard і отримувати відповіді, з можливістю обробляти затримки і помилки,

встановлюючи затримки. Однак цей процес може бути громіздким, оскільки він передбачає роботу з файлами cookie і потенційне використання проксі-серверів, щоб уникнути блокування або необхідності розв'язувати CAPTCHA. Проксі-сервери можна налаштувати на ротацію IP-адрес, що допомагає обходити CAPTCHA, але це додає ще один рівень складності до налаштування. Сумуючи вище описане, хоча Bard API надає метод взаємодії з Google Bard, необхідність налаштування файлів cookie, їх часті зміни, введення CAPTCHA та потенційного налаштування проксі-серверів робить цей метод менш зручним для постійного використання.

Бібліотека `gpt4free` [68], являє собою колекцію з відкритим вихідним кодом, яка пропонує доступ до кількох найсучасніших мовних моделей. Вона підтримує не лише GPT-4 і GPT-3.5, але й дозволяє використовувати різні провайдери, які слугують інтерфейсами до цих моделей. До складу бібліотеки входять такі провайдери:

- `Bing` (`g4f.Provider.Bing`): Пропонує GPT-4 з можливостями потокового передавання, але не вимагає автентифікації;
- `GeekGpt` (`g4f.Provider.GeekGpt`): Підтримує як GPT-3.5, так і GPT-4 з можливістю потокового передавання;
- `GptChatly` (`g4f.Provider.GptChatly`): Надає доступ до GPT-3.5 і GPT-4, але без підтримки потокового передавання;
- `Liaobots` (`g4f.Provider.Liaobots`): Підтримує GPT-3.5 і GPT-4, а також потокове передавання;
- `Phind` (`g4f.Provider.Phind`): Підтримує лише GPT-4 з функціями потокового передавання;
- `Raaycast` (`g4f.Provider.Raycast`): Пропонує як GPT-3.5, так і GPT-4 з потоковою передачею і вимагає автентифікації.

Ці провайдери є джерелами, з яких бібліотека `gpt4free` може отримувати та виконувати відповіді мовної моделі. Бібліотека інкапсулює складність взаємодії з різними API, представляючи уніфікований і спрощений інтерфейс, що дозволяє включити розширену обробку мови III у свої системи.

`gpt4free` відрізняється від GPT4All і Bard API тим, що пропонує безпроблемний процес встановлення, широкий спектр потужних мовних моделей і простий підхід

доступу до різних постачальників моделей. Це робить gpt4free дуже доступним і ефективним вибором для реалізації поставленого в роботі завдання.

Обравши модель можна сформулювати необхідні завдання які потрібно вирішити в ході програмної реалізації gpt4free:

1. Розробити алгоритм за яким бібліотека буде обробляти необхідний запит та видавати коректну відповідь;
2. Розробити зручний та інтуїтивний інтерфейс користувача для демонстрації роботи модуля;
3. Опрацювати відповідь від моделі та виводити її в зручному та правильному форматі;
4. Провести тестування та оцінити результати розробленого програмного модуля.

3.2 Структура і функціонал розробленого модуля автоматизованої системи

Для програмної реалізації модуля було використано наступні бібліотеки:

- Бібліотека streamlit використовується для розробки інтерактивних веб-додатків та візуалізації даних. Вона надає простий спосіб створення веб-інтерфейсу;
- Модуль os надає функціональність, пов'язану з операційною системою. Він дозволяє взаємодіяти з файловою системою, отримувати інформацію про середовище виконання програми, запускати команди оболонки та інші операції, пов'язані з ОС;
- Модуль datetime надає функціональність для роботи з датами та часом. Він дозволяє отримувати поточну дату та час, виконувати операції з датами, формувати та розбирати рядки дати і часу;
- PyPDF2: Бібліотека PyPDF2 використовується для роботи з файлами у форматі PDF. Вона надає можливості для читання, запису та редагування PDF-файлів, таких як видобування тексту, об'єднання та розділення сторінок, налаштування захисту і багато іншого;

– g4f бібліотека для використання моделі ШІ.

Основною функцією в модулі є `summarize_text` призначена для здійснення автоматичного аналізу, але перед її аналізом потрібно розглянути допоміжну функцію `main`. (рис. 3.2).

```
def main():
    st.title("Додаток для формування короткого фінансового звіту компанії PDF-тексту")
    uploaded_file = st.file_uploader("Завантажити PDF-файл", type=["pdf"])
    if uploaded_file:
        st.success("Файл успішно завантажено!")
        with st.spinner("Аналіз та формування короткого фінансового звіту..."):
            filename = uploaded_file.name
            base_name = os.path.basename(filename)
            name_without_extension = os.path.splitext(base_name)[0]
            pdfFileObject = open(filename, 'rb')
            pdfReader = PdfReader(pdfFileObject)
            text = []
            for i in range(0, len(pdfReader.pages)):
                pageObj = pdfReader.pages[i].extract_text()
                pageObj = pageObj.replace('\t\r', '')
                pageObj = pageObj.replace('\xa0', '')
                text.append(pageObj)
            new_text = join_elements(text, 3)
            new_text_len = len(new_text)
            summary = summarize_text(new_text)
            st.subheader("Результат")
            st.write(summary)
            with open(f'./{name_without_extension}_summary_ui.txt', 'w') as out:
                out.write(summary)
```

Рисунок 3.2 — Лістинг функції `main`

Код створює короткий фінансовий звіт компанії з PDF-тексту. Спочатку він відображає заголовок сторінки і створює поле для завантаження PDF-файлу. Після перевірки завантаження файлу та відображення повідомлення про успіх, відбувається аналіз файлу. Код витягує текст з кожної сторінки PDF, обробляє його, видаляє зайві символи, додає текст до списку, а потім об'єднує елементи списку в новий текст. Завершується це підсумуванням тексту за допомогою функції `summarize_text` для формування звіту.

Функція `summarize_text` (рис. 3.3) призначена для здійснення автоматичного аналізу тексту. Вхідним параметром функції є `text`, який містить список абзаців тексту. У циклі `for` проходимо по кожному елементу у списку `text`. Створюємо змінну `prompt`, яка містить текстову інструкцію для моделі, що пояснює її завдання. Це завдання полягає у аналізі фінансового звіту, тобто тексту. Запитуємо модель,

використовуючи функцію `get_completion`, передаючи їй `prompt`. Отримуємо відповідь моделі, яка містить підсумок тексту. Додаємо підсумок до змінної `summary_chunks`. Чекаємо 5 секунд, перш ніж перейти до наступного абзацу тексту, використовуючи `time.sleep(5)`.

```
def summarize_text(text):
    summary_chunks = ''
    for i in range(len(text)):
        prompt = f"""
            Your task is to act as a Text Summariser.
            I'll give you text from pages of a book from beginning to end.
            And your job is to summarise text from these pages.
            Don't be conversational. I need a plain answer.
            Text is shared below, delimited with triple backticks:
            ```{text[i]}```
 """
 try:
 response = get_completion(prompt)
 except:
 response = get_completion(prompt)
 summary_chunks = summary_chunks + ' ' + response + '\n\n'
 time.sleep(5)
 main_prompt = f"""
 Your task is to act as a text summarizer.
 I will provide you with text from pages of a book from beginning to end,
 and your job is to summarize and provide an assessment of their financial position the text from these pages.
 Do not be conversational. I need a plain answer.
 Begin with "Короткий фінансовий звіту компанії {company_name}
 (Replace with the name of the company from the text is shared below, delimited with triple backticks):
 The text is shared below, delimited with triple backticks:
    ```{summary_chunks}```
    The result will be translated into the Ukrainian language.
    """
    try:
        summary = get_completion(main_prompt)
    except:
        summary = get_completion(main_prompt)
    st.success('Програма успішно завершила свою роботу!')
    return summary
```

Рисунок 3.3 — Лістинг функції `summarize_text`

Після завершення циклу, створюємо змінну `main_prompt`, яка містить загальну інструкцію для моделі щодо підсумовування тексту. Знову запитуємо модель, використовуючи функцію `get_completion`, передаючи їй `main_prompt`. Отримуємо відповідь моделі, яка містить результат підсумків тексту зі всіх абзаців. Повертаємо підсумок з функції. Ця функція використовується для автоматичного аналізу тексту, розділеного на абзаци. Вона спирається на модель `gpt-3.5-turbo`, щоб генерувати підсумки на основі наданих інструкцій. Процес триває деякий час через затримки у виконанні та очікування моделі. Коли функція завершується, повертається загальний підсумок підсумків тексту.

Розглянемо ключові запити до моделі `prompt` (рис. 3.4) та `main_prompt` детальніше адже вони безпосередньо взаємодіють з моделлю.

```
prompt = f"""
Your task is to act as a Text Summariser.
I'll give you text from pages of a book from beginning to end.
And your job is to summarise text from these pages.
Don't be conversational. I need a plain answer.
Text is shared below, delimited with triple backticks:
```{text[i]}```
"""
```

Рисунок 3.4 — Лістинг запиту prompt

У цьому конкретному запиті модель отримує наступне завдання:

- Бути текстовим аналізатором ( підсумувати текст);
- Отримувати текст зі сторінок книги, починаючи з початку і закінчуючи в кінці;
- Завдання полягає у підсумовуванні тексту з цих сторінок;
- Не бути розмовною. Потрібна чітка відповідь;
- Текст надається нижче, відокремлений абзацами потрійними зворотними апострофами.

main\_prompt конкретизує завдання (рис. 3.5), та перекладає його на українську мову. Запити створюються англійською адже модель краще розуміє її та швидше обробляє запит.

```
main_prompt = f"""
Your task is to act as a text summarizer.
I will provide you with text from pages of a book from beginning to end,
and your job is to summarize and provide an assessment of their financial position the text from these pages.
Do not be conversational. I need a plain answer.
Begin with "Короткий фінансовий звіт компанії company_name
(Replace with the name of the company from the text is shared below, delimited with triple backticks):
The text is shared below, delimited with triple backticks:
```{summary_chunks}```
The result will be translated into the Ukrainian language.
"""
```

Рисунок 3.5 — Лістинг запиту main_prompt

В основі цього запиту, наступні завдання:

- Бути текстовим аналізатором (підсумувати текст);

- Отримувати текст зі сторінок книги, починаючи з початку і закінчуючи в кінці;
- Завдання полягає у підсумовуванні тексту з цих сторінок та наданні оцінки фінансового стану компанії, яка згадується у тексті.
- Не бути розмовною. Потрібна чітка відповідь;
- Почати підсумок з "Короткого фінансового звіту компанії `company_name` (Замініть на назву компанії з тексту, який надається нижче, відокремлений потроху потрійними зворотними апострофами);
- Текст надається нижче, відокремлений потроху потрійними зворотними апострофами;
- Результат перекласти на українську мову.

Виконана програмна реалізація дозволяє модулю виконувати аналіз будь якого фінансового звіту за будь яким шаблоном та формуванням звітності. Результат аналізу також записується в текстовий документ для зручності. Враховуючи написання модулю на мові Python його з легкістю можливо інтегрувати у вже існуючу систему, що робить його адаптивним та гнучким у використанні.

У розроблений модуль інтегрований інтерфейс (рис. 3.6), який дозволяє відобразити реакцію коду на дії користувача та дає користувачу необхідну інформацію про процес аналізу. Для перевірки роботи модулю та його ефективності було обрано фінансові звіти компаній:

1. Management's Discussion and Analysis of Financial Condition and Results of Operations (Microsoft)[69];
2. Alphabet Inc. Consolidated balance Sheet.[70];
3. Apple Inc. CONDENSED CONSOLIDATED STATEMENTS OF OPERATIONS (Unaudited) [71].

Коли модуль запущено користувач може обрати звіт та завантажити для аналізу (рис. 3.6) модуль надасть йому повідомлення про це.

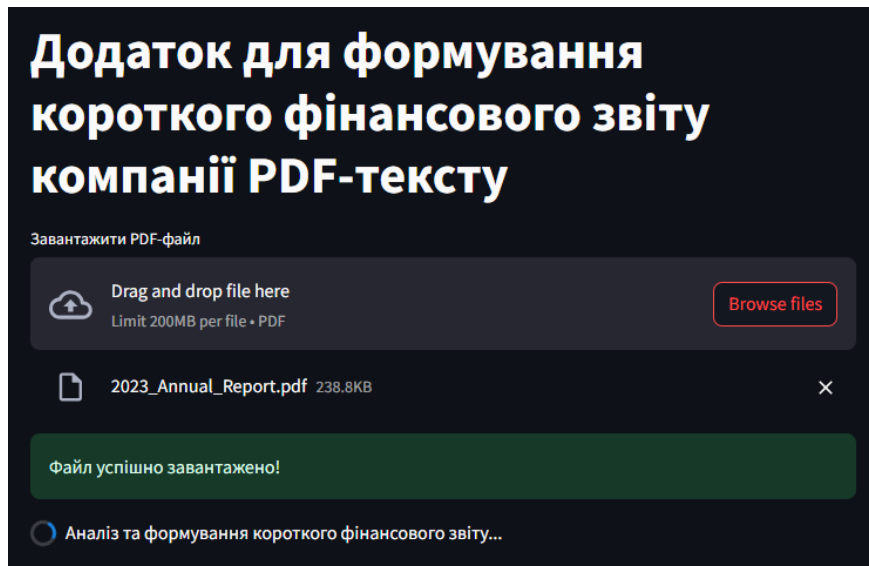


Рисунок 3.6 — Інтерфейс модулю та повідомлення

Проаналізувавши фінансові звіти компаній Microsoft, Alphabet, Apple було отримано хороший результат у вигляді тексту який структуровано та чітко описує ключові моменти фінансового звіту. Не дивлячись на різноманіття отриманої інформації її тип та об'єм модуль виконує свої функції швидко та ефективно.

Результати аналізу та демонстрація сторінок звіту компаній наведені в додатках.

3.3 Оцінка очікуваного ефекту від впровадження модуля для автоматизованого аналізу фінансового звіту за допомогою машинного навчання

Оцінка розробки та впровадження модуля для автоматизованого аналізу фінансових звітів за допомогою машинного навчання це важливий етап який показує економічну доцільність даної розробки. Модуль, інтегрований в існуючі системи, спрямований на оптимізацію процесу аналізу, скорочення часу та зусиль, необхідних фінансовим аналітикам.

Насамперед потрібно провести розрахунок капітальних витрат, який являє собою процес визначення загальної суми грошей, яку потрібно вкласти для створення та впровадження проекту. Стосовно модулю то тут потрібно врахувати 2 ключові витрати:

- заробітна плата розробника за місяць;
- оплата API для використання штучного інтелекту GPT за рік часу.

За даними DOU [72] середня зарплата розробника Python складає 2700 доларів (рис. 3.7) або 98 782,58 гривень.

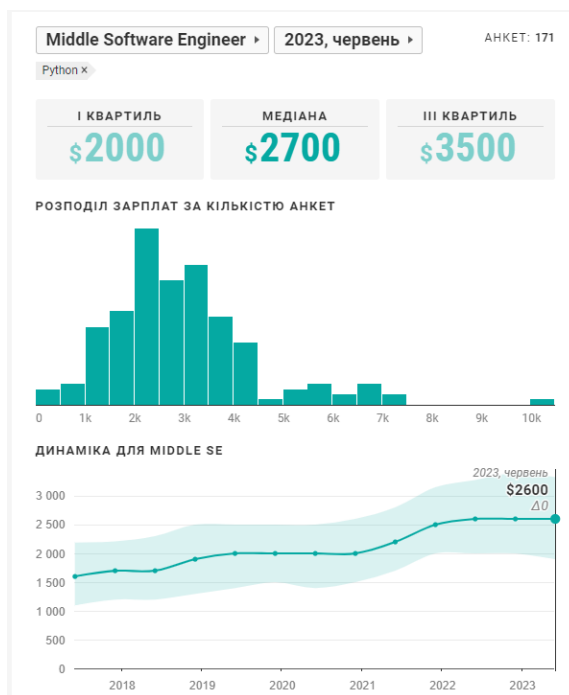


Рисунок 3.7 — Середня заробітна плата Middle Software Engineer

Що до API GPT воно обійдеться в 18 доларів або 658,55 гривень. Сумуючи вище отриманні значення розрахуємо капітальні витрати (S) використовуючи формулу 3.1:

$$S = V_{zn} + (V_{ка} \cdot K_m) \quad (3.1)$$

де:

V_{zn} – Витрати заробітної плати розробника;

$V_{ка}$ – витрати на використання API ключа;

K_m – Кількість місяців використання API ключа.

Отже капітальні витрати модулю становлять 106,685 гривень. Щоб розрахувати скільки можливо буде заощадити (R) використовуючи модуль скористаємось формулою 3.2:

$$R = (P_a \cdot K_m) - S \quad (3.2)$$

де:

P_a – Ручний аналіз фінансових звітів;

K_m – Кількість місяців;

S – Капітальні витрати на реалізацію модулю.

За даними Work.ua [73] середня зарплата фінансиста становить 27500 на місяць. Для ручного аналізу фінансового звіту об'ємом в 50-100 сторінок необхідно буде в середньому місяць робочого часу. Таким чином використовуючи ці дані, 10 робочих місяців, та формулу 3.2 отримуємо значення 168,315 гривень які можливо заощадити використовуючи розроблений модуль.

Період окупності використовуючи модуль можна розрахувати як співвідношення капітальних витрат до річної економії використовуючи формулу 3.3:

$$P = \frac{S}{R} \quad (3.3)$$

де:

S – Капітальні витрати на реалізацію модулю;

R – Річна економія, використовуючи розроблений модуль.

Отже, період окупності буде 232 дні. Це час, який знадобиться для того, щоб початкові інвестиції були повністю повернуті за рахунок заощаджень, отриманих в результаті реалізації проекту. Впровадження автоматизованого модуля для аналізу фінансових звітів демонструє швидкий період окупності, що підкреслює його економічну ефективність та потенціал для значної щорічної економії. Це робить його фінансово життєздатним проектом, що пропонує значне підвищення ефективності порівняно з традиційними ручними методами аналізу.

ВИСНОВОК

У кваліфікаційній роботі досліджено проблеми генерації текстових даних з фінансових звітів. Описано загальні характеристики цього процесу, підкреслено важливість аналізу фінансових звітів для генерації текстових даних.

Наступна частина роботи присвячена технологіям, що використовуються для вирішення завдання генерації текстових даних. Оглянуто загальні методи машинного навчання. Проаналізовано конкретні технології генерації тексту з використанням нейронних мереж. Досліджено потенціал та перспективи машинного навчання у створенні текстових даних на основі аналізу фінансових звітів.

Далі було детально описано процес розробки модуля, призначеного для автоматизованого аналізу фінансових звітів. Створено технічне завдання, яке необхідне для вирішення поставленої задачі. Потім описано структуру та функціональність розробленого модуля в автоматизованій системі. Нарешті, представлено оцінку очікуваного ефекту від впровадження модуля з використанням машинного навчання для автоматизованого аналізу фінансової звітності.

Таким чином, кваліфікаційна робота включає в себе комплексне дослідження проблеми генерації текстових даних з фінансових звітів. У дослідженні розглядаються різні аспекти, такі як загальна характеристика, аналіз існуючих сервісів та можливості, що надаються методами машинного навчання. Результатом виконання роботи є розроблений модуль для автоматизованого аналізу фінансових звітів. Враховуючи те що модуль створений за допомогою Python він може бути легко інтегрований в різні системи або функціонувати як незалежна система, наприклад, веб-сайт або програма для персонального комп'ютера.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Конституція України (Відомості Верховної Ради України (ВВР), 1996, № 30, ст. 141). <https://zakon.rada.gov.ua/laws/show/254к/96-вр#Text> (дата звернення: 06.02.2023)
2. Генеративні мережі. URL: <https://www.victoria.lviv.ua/library/students/sss/theme10.html> (дата звернення: 29.10.2023)
3. Копей, В. (2021). Простий генератор інтерактивних документів juryter notebook з розміченого програмного коду PYTHON. Вісник Університету «Україна» Серія Інформатика, обчислювальна техніка та кібернетика, 1(22). вилучено із <https://visn-it.uu.edu.ua/index.php/visn-icct/article/view/31> (дата звернення: 29.10.2023)
4. Платформа @Rev. URL: <https://www.rev.com> (дата звернення: 29.10.2023)
5. Платформа Descript. URL: https://www.descript.com/?lmref=P_awkw&gclid=Cj0KCQjwhfipBhCqARIsAH9msbk1D0eJqMDWPo8ryEHy1Xb1nOQRx5EyZyw9l2aylMwZ4KI5gTchO9UaAsSvEALw_wcB (дата звернення: 29.10.2023)
6. Платформа LangPie. URL: <https://lingopie.com> (дата звернення: 29.10.2023)
7. Платформа Lofty AI. URL: <https://www.lofty.ai> (дата звернення: 29.10.2023)
8. Платформа Otter. URL: <https://otter.ai> (дата звернення: 29.10.2023)
9. Платформа Speak Ukrainian. URL: <https://www.speakua.com> (дата звернення: 29.10.2023)
10. Поняття про бази даних. URL: <https://naurok.com.ua/test/ponyattya-pro-bazi-danih-1961289.html> (дата звернення: 29.10.2023)
11. Потапов Д.С. Дослідження нейромережевих методів зіставлення зображень та їх текстових анотацій. Кваліфікаційна робота. 2022 рік. 78 с. URL: <https://openarchive.nure.ua/server/api/core/bitstreams/554d9014-6e0f-4e02-a039-80260f4b340d/content> (дата звернення: 29.10.2023)

12. Юсин, Я. О. Генерування корпусів текстових даних на основі детермінованого методу / Я. О. Юсин, Т. М. Заболотня // Наукові вісті КПП: міжнародний науково-технічний журнал. – 2021. – № 3(133). – С. 38–45.
13. Amazon Transcribe. URL: <https://docs.aws.amazon.com/managedservices/latest/userguide/transcribe.html> (дата звернення: 29.10.2023)
14. Google Cloud Speech-to-Text. URL: <https://cloud.google.com/speech-to-text> (дата звернення: 29.10.2023)
15. Платформа IBM Watson Speech to Text. URL: <https://www.ibm.com/products/speech-to-text#:~:text=What%20is%20IBM%20Watson%20Speech,agent%20assistance%20and%20speech%20analytics>. (дата звернення: 29.10.2023)
16. III та генерація тексту в зображення. URL: <https://ts2.space/uk/iii-ta-generacija-tekstu-v-zobrazennja/>(дата звернення: 29.10.2023)
17. Платформа Quill | Narrative Science. URL: <https://www.cabinetm.com/product/narrative-science/quill> (дата звернення: 29.10.2023)
18. Платформа Yseop. URL: <https://yseop.com> (дата звернення: 29.10.2023)
19. Платформа Tableau Software Training and Tutorials. URL: https://www.pluralsight.com/browse/data-professional/tableau?utm_source=google&utm_medium=paid-search&utm_campaign=upskilling-and-reskilling&utm_term=ssi-emea-dynamic&utm_content=free-trial&gad_source=1&gclid=Cj0KCQiAr8eqBhD3ARIsAie-buNjiDWv4kIQi2EYk2ZEmPN12M86DPbqSUEjdN6Y1bF6ZuPNQks37ZMaAoe4EALw_wcB (дата звернення: 29.10.2023)
20. Платформа Microsoft Power BI - Analiza Danych Biznesowych BI. URL: <https://powerbi.microsoft.com/ru-ru/> (дата звернення: 29.10.2023)
21. Платформа ClickAI. URL: https://clickai.com/welcome_to_clickai (дата звернення: 29.10.2023)
22. IBM Watson. URL: <https://www.ibm.com/watson> (дата звернення: 29.10.2023)

23. Google Cloud Natural Language. URL: <https://cloud.google.com/natural-language/docs> (дата звернення: 29.10.2023)

24. Google Cloud BigQuery - Google Cloud Platform (GCP). URL: https://cloud.google.com/bigquery?utm_source=google&utm_medium=cpc&utm_campaign=emea-pl-all-en-dr-bkws-all-all-trial-e-gcp-1011340&utm_content=text-ad-none-any-dev_c-cre_574628703097-adgp_Hybrid+%7C+BKWS+-+EXA+%7C+Txt+~+Data+Analytics+~+BigQuery%23v1-kwid_43700072692462204-aud-606988877934:kwd-372661971564-userloc_1011517&utm_term=kw_bigquery-net_g-plac_&&gad_source=1&gclid=Cj0KCQiAr8eqBhD3ARIsAie-buM29xhlCKCt87Rpb52sRkPYiKuSUF2InmU4mmx2EC6V_CuE_IqwV-4aAtDJEALw_wcB&gclsrc=aw.ds (дата звернення: 29.10.2023)

25. AutoML. URL: <https://cloud.google.com/automl> (дата звернення: 29.10.2023)

26. Microsoft Excel. URL: <https://www.microsoft.com/pl-pl/microsoft-365/excel> (дата звернення: 29.10.2023)

27. Microsoft Power Query. URL: <https://powerquery.microsoft.com/en-us/> (дата звернення: 29.10.2023)

28. Надстройка для Excel - Power Pivot. URL: <https://support.microsoft.com/ru-ru/office/power-pivot-обзор-и-обучение-f9001958-7901-4caa-ad80-028a6d2432ed> (дата звернення: 29.10.2023)

29. Штучний інтелект, машинне навчання та нейронні мережі: в чому різниця і для чого їх використовують. URL: <https://evergreens.com.ua/ua/articles/machine-learning-overview.html> (дата звернення: 29.10.2023)

30. Машинне навчання простими словами. Частина 1. URL: <http://www.mmf.lnu.edu.ua/en/aren/1739> (дата звернення: 29.10.2023)

31. Semi-Supervised Learning: Techniques & Examples [2023]. URL: <https://www.v7labs.com/blog/semi-supervised-learning-guide#:~:text=we%27ve%20covered%3A-,Semi-supervised%20learning%20is%20a%20broad%20category%20of%20machine%20learning,save%20valuable%20time%20and%20money.> (дата звернення: 29.10.2023)

32. The Beginner's Guide to Deep Reinforcement Learning. URL: <https://www.v7labs.com/blog/deep-reinforcement-learning-guide> (дата звернення: 29.10.2023)

33. The Beginner's Guide to Self-Supervised Learning. URL: <https://www.v7labs.com/blog/self-supervised-learning-guide> (дата звернення: 29.10.2023)

34. What is unsupervised learning? URL: <https://cloud.google.com/discover/what-is-unsupervised-learning#:~:text=Unsupervised%20learning%20in%20artificial%20intelligence,any%20explicit%20guidance%20or%20instruction.> (дата звернення: 29.10.2023)

35. Створення розумнішого штучного інтелекту: вичерпний огляд алгоритмів і методів інтелектуального аналізу даних. URL: <https://ts2.space/uk/створення-розумнішого-штучного-інте-8/> (дата звернення: 29.10.2023)

36. What is Deep Learning? URL: <https://www.ibm.com/topics/deep-learning> (дата звернення: 29.10.2023)

37. Що таке CNN (згорточні нейронні мережі)? <https://www.unite.ai/uk/what-are-convolutional-neural-networks/> (дата звернення: 29.10.2023)

38. Анатомія RNN: аналіз компонентів рекурентної нейронної мережі. URL: <https://ts2.space/uk/анатомія-rnn-аналіз-компонентів-рекурен/> (дата звернення: 29.10.2023)

39. Мітап «CPU, GPU, TPU, PPU. Яке це має відношення до хмарних технологій?» (подія в архіві). URL: <https://dou.ua/calendar/38628/> (дата звернення: 29.10.2023)

40. Трохановський В. І. Ансамблеві методи класифікації в машинному навчанні [Електронний ресурс] / Трохановський В. І. // Цифрова економіка: зб. матеріалів II Нац. наук.-метод. конф., 17–18 жовт. 2019 р., м. Київ / М-во освіти і науки України, ДВНЗ «Київ. нац. екон. ун-т ім. В. Гетьмана» [та ін.]; [оргком.: Д. Г. Лук'яненко (голова) та ін.]. – Електрон. текст. дані. – Київ: КНЕУ, 2019. – С. 41–43.

41. What is Random Forest? | IBM. URL: <https://www.ibm.com/topics/random-forest#:~:text=Random%20forest%20is%20a%20commonly,both%20classification%20and%20regression%20problems>. (дата звернення: 29.10.2023)
42. A Gentle Introduction to the Gradient Boosting Algorithm for Machine Learning. URL: <https://machinelearningmastery.com/gentle-introduction-gradient-boosting-algorithm-machine-learning/> (дата звернення: 29.10.2023)
43. What is random forest? URL: <https://www.ibm.com/topics/random-forest#:~:text=Random%20forest%20is%20a%20commonly,both%20classification%20and%20regression%20problems>. (дата звернення: 29.10.2023)
44. All You Need to Know about Gradient Boosting Algorithm – Part 1. Regression. URL: <https://towardsdatascience.com/all-you-need-to-know-about-gradient-boosting-algorithm-part-1-regression-2520a34a502> (дата звернення: 29.10.2023)
45. Introduction to Recurrent Neural Network. URL: <https://www.geeksforgeeks.org/introduction-to-recurrent-neural-network/> (дата звернення: 29.10.2023)
46. A Gentle Introduction to Long Short-Term Memory Networks by the Experts. URL: <https://machinelearningmastery.com/gentle-introduction-long-short-term-memory-networks-experts/> (дата звернення: 29.10.2023)
47. What are Recurrent Neural Networks? URL: <https://www.ibm.com/topics/recurrent-neural-networks> (дата звернення: 29.10.2023)
48. How Transformers Work. URL: <https://towardsdatascience.com/transformers-141e32e69591> (дата звернення: 29.10.2023)
49. Attention Mechanism in Deep Learning. URL: <https://www.scaler.com/topics/deep-learning/attention-mechanism-deep-learning/#:~:text=Attention%20mechanisms%20are%20a%20type,to%20the%20task%20at%20hand>. (дата звернення: 29.10.2023)
50. GPT (Generative Pre-trained Transformer) від OpenAI. URL: <https://openai.com> (дата звернення: 29.10.2023)

51. BERT: Bidirectional Encoder Representations from Transformers – Unlocking the Power of Deep Contextualized Word Embeddings. URL: <https://learnopencv.com/bert-bidirectional-encoder-representations-from-transformers/> (дата звернення: 29.10.2023)
52. GPT-4. URL: <https://openai.com/research/gpt-4> (дата звернення: 29.10.2023)
53. GPT-3 powers the next generation of apps. URL: <https://openai.com/product> (дата звернення: 29.10.2023)
54. RoBERTa: A Robustly Optimized BERT Pretraining Approach. URL: <https://www.laellerr.com/blog/roberta-a-robustly-optimized-bert-pretraining-approach/> (дата звернення: 29.10.2023)
55. AI DistilBERT: Революціонізація обробки природних мов. URL: <https://ts2.space/uk/ai-distilbert-революціонізація-обробки-природн/> (дата звернення: 29.10.2023)
56. Sequence-to-Sequence Architecture Made Easy & How To Tutorial In Python. URL: <https://spotintelligence.com/2023/09/28/sequence-to-sequence/> (дата звернення: 29.10.2023)
57. Conditional Random Fields (CRFs). URL: <https://www.geeksforgeeks.org/conditional-random-fields-crfs-for-pos-tagging-in-nlp/> (дата звернення: 29.10.2023)
58. What Is Named Entity Recognition (NER)? URL: <https://www.techtarget.com/whatis/definition/named-entity-recognition-NER> (дата звернення: 29.10.2023)
59. Autoencoders in Deep Learning: Tutorial & Use Cases [2023]. URL: <https://www.v7labs.com/blog/autoencoders-guide> (дата звернення: 29.10.2023)
60. What is generative AI? Everything you need to know. URL: <https://www.techtarget.com/searchenterpriseai/definition/variational-autoencoder-VAE> (дата звернення: 29.10.2023)
61. Attention Mechanism in Deep Learning- Scaler Topics. URL: <https://www.scaler.com/topics/deep-learning/attention-mechanism-deep-learning/> (дата звернення: 29.10.2023)

62. Гаврик А. А. ChatGPT як інноваційний помічник в управлінні фінансовою стійкістю / А. А. Гаврик, Т. Ю. Назарова // Вісник Національного технічного університету "ХПІ" (економічні науки) = Bulletin of the National Technical University "KhPI" (economic sciences): зб. наук. пр. – Харків: НТУ "ХПІ", 2023. – № 3. – С. 47-51.
63. DOU. Генерація текстів: перевіряємо прогрес AI-моделі від GPT до ChatGPT. URL: <https://dou.ua/forums/topic/41509/>.
64. OpenAI. OpenAI API. URL: <https://openai.com/blog/openai-api>.
65. OpenAI. Pricing. URL: <https://openai.com/pricing>.
66. nomic-ai. GPT4All. URL: <https://github.com/nomic-ai/gpt4all>.
67. dsdanielpark. Bard-API. URL: <https://github.com/dsdanielpark/Bard-API>.
68. xtekky. gpt4free. URL: <https://github.com/xtekky/gpt4free>.
69. Annual Report. Management's Discussion and Analysis of Financial Condition and Results of Operations. URL: microsoft.com/investor/reports/ar13/financial-review/discussion-analysis/index.html;
70. Alphabet Inc. Consolidated balance Sheet. URL: sec.gov/Archives/edgar/data/1652044/000165204423000016/goog-20221231.htm;
71. Apple Inc. CONDENSED CONSOLIDATED STATEMENTS URL: apple.com/newsroom/pdfs/FY23_Q2_Consolidated_Financial_Statements.pdf;
72. Зарплати Middle Python розробників URL: jobs.dou.ua/salaries/?period=2023-06&position=Middle%20SE&technology=Python;
73. Фінансист: середня зарплата в Україні URL: <https://www.work.ua/salary-%D1%84%D1%96%D0%BD%D0%B0%D0%BD%D1%81%D0%B8%D1%81%D1%82/>

ДОДАТКИ

Додаток А. Аналіз фінансового звіту Microsoft

Сторінка фінансового звіту:

LinkedIn revenue growth	Revenue from LinkedIn, including Talent Solutions, Marketing Solutions, Premium Subscriptions, and Sales Solutions
Server products and cloud services revenue growth	Revenue from Server products and cloud services, including Azure and other cloud services; SQL Server, Windows Server, Visual Studio, System Center, and related Client Access Licenses ("CALs"); and Nuance and GitHub

More Personal Computing

Metrics related to our More Personal Computing segment assess the performance of key lines of business within this segment. These metrics provide strategic product insights which allow us to assess the performance across our commercial and consumer businesses. As we have diversity of target audiences and sales motions within the Windows business, we monitor metrics that are reflective of those varying motions.

Windows OEM revenue growth	Revenue from sales of Windows Pro and non-Pro licenses sold through the OEM channel
Windows Commercial products and cloud services revenue growth	Revenue from Windows Commercial products and cloud services, comprising volume licensing of the Windows operating system, Windows cloud services, and other Windows commercial offerings
Devices revenue growth	Revenue from Devices, including Surface, HoloLens, and PC accessories
Xbox content and services revenue growth	Revenue from Xbox content and services, comprising first- and third-party content (including games and in-game content), Xbox Game Pass and other subscriptions, Xbox Cloud Gaming, advertising, third-party disc royalties, and other cloud services
Search and news advertising revenue (ex TAC) growth	Revenue from search and news advertising excluding traffic acquisition costs ("TAC") paid to Bing Ads network publishers and news partners

SUMMARY RESULTS OF OPERATIONS

<u>(In millions, except percentages and per share amounts)</u>	<u>2023</u>	<u>2022</u>	<u>Percentage Change</u>
Revenue	\$ 211,915	\$ 198,270	7%
Gross margin	146,052	135,620	8%
Operating income	88,523	83,383	6%
Net income	72,361	72,738	(1)%
Diluted earnings per share	9.68	9.65	0%
Adjusted gross margin (non-GAAP)	146,204	135,620	8%
Adjusted operating income (non-GAAP)	89,694	83,383	8%
Adjusted net income (non-GAAP)	73,307	69,447	6%
Adjusted diluted earnings per share (non-GAAP)	9.81	9.21	7%

Adjusted gross margin, operating income, net income, and diluted earnings per share ("EPS") are non-GAAP financial measures. Current year non-GAAP financial measures exclude the impact of the Q2 charge, which includes employee severance expenses, impairment charges resulting from changes to our hardware portfolio, and costs related to lease consolidation activities. Prior year non-GAAP financial measures exclude the net income tax benefit related to transfer of intangible properties in the first quarter of fiscal year 2022. Refer to Note 12 – Income Taxes of the Notes to Financial Statements for further discussion. Refer to the Non-GAAP Financial Measures section below for a reconciliation of our financial results reported in accordance with GAAP to non-GAAP financial results.

Результат:

Короткий фінансовий звіт компанії Microsoft Corporation:

- В фінансовому році 2023 року дохід компанії збільшився на 13,6 мільярда доларів або 7% порівняно з 2022 роком.
- Валовий прибуток збільшився на 10,4 мільярда доларів або 8%.
- Операційний прибуток збільшився на 5,1 мільярда доларів або 6%.
- Чистий прибуток зменшився на 1%.
- Розведений прибуток на акцію залишився незмінним і становив 9,68 долара.
- Скоригований валовий прибуток, операційний прибуток, чистий прибуток та розведений прибуток на акцію (непо GAAP) також збільшилися.
- Зростання доходу було зумовлено зростанням в галузі Інтелектуального хмарного обчислення та Продуктивності та бізнес-процесів, але зменшенням в галузі Особистого комп'ютера.
- Доходи від Windows OEM та пристроїв зменшилися.
- Доходи від контенту та послуг Xbox також зменшилися.
- Доходи від реклами пошуку та новин (за винятком витрат на привертання трафіку) збільшилися.
- Фінансові результати включали негативний вплив зміни валютного курсу.

Загалом, компанія Microsoft зазнала змішаних результатів у фінансовому році 2023 року. Вона зазнала зростання доходу в деяких сегментах, таких як Інтелектуальне хмарне обчислення та Продуктивність та бізнес-процеси, але зазнала зниження в інших сегментах, таких як Особистий комп'ютер. Зменшення доходу від Windows OEM та пристроїв, а також від контенту та послуг Xbox, може бути причиною зниження чистого прибутку компанії. Крім того, негативний вплив зміни валютного курсу також вплинув на фінансові результати компанії.

Будь ласка, зверніть увагу, що це короткий виклад змісту наданого тексту і може не містити всіх деталей з оригінального джерела.

Додаток Б. Аналіз фінансового звіту Alphabet Inc

Сторінка фінансового звіту:

PART I. FINANCIAL INFORMATION
ITEM 1. FINANCIAL STATEMENTS

Alphabet Inc.
CONSOLIDATED BALANCE SHEETS
(in millions, except par value per share amounts)

	As of December 31, 2022	As of September 30, 2023 (unaudited)
Assets		
Current assets:		
Cash and cash equivalents	\$ 21,879	\$ 30,702
Marketable securities	91,883	89,233
Total cash, cash equivalents, and marketable securities	113,762	119,935
Accounts receivable, net	40,258	41,020
Inventory	2,670	2,957
Other current assets	8,105	12,398
Total current assets	164,795	176,310
Non-marketable securities	30,492	30,907
Deferred income taxes	5,261	10,983
Property and equipment, net	112,668	125,705
Operating lease assets	14,381	14,199
Intangible assets, net	2,084	1,833
Goodwill	28,960	29,146
Other non-current assets	6,623	7,628
Total assets	<u>\$ 365,264</u>	<u>\$ 396,711</u>
Liabilities and Stockholders' Equity		
Current liabilities:		
Accounts payable	\$ 5,128	\$ 5,803
Accrued compensation and benefits	14,028	12,562
Accrued expenses and other current liabilities	37,866	55,602
Accrued revenue share	8,370	8,025
Deferred revenue	3,908	4,303
Total current liabilities	69,300	86,295
Long-term debt	14,701	13,781
Deferred revenue, non-current	599	884
Income taxes payable, non-current	9,258	8,038
Deferred income taxes	514	528
Operating lease liabilities	12,501	12,550
Other long-term liabilities	2,247	1,433
Total liabilities	<u>109,120</u>	<u>123,509</u>
Commitments and contingencies (Note 9)		
Stockholders' equity:		
Preferred stock, \$0.001 par value per share, 100 shares authorized; no shares issued and outstanding	0	0
Class A, Class B, and Class C stock and additional paid-in capital, \$0.001 par value per share: 300,000 shares authorized (Class A 180,000, Class B 60,000, Class C 60,000); 12,849 (Class A 5,964, Class B 883, Class C 6,002) and 12,541 (Class A 5,922, Class B 874, Class C 5,745) shares issued and outstanding	68,184	74,591
Accumulated other comprehensive income (loss)	(7,603)	(7,036)
Retained earnings	195,563	205,647
Total stockholders' equity	<u>256,144</u>	<u>273,202</u>
Total liabilities and stockholders' equity	<u>\$ 365,264</u>	<u>\$ 396,711</u>

See accompanying notes.

Результат:

Короткий фінансовий звіт компанії Google:

За останні 3 роки загальний дохід Google збільшився з 135,2 мільярда доларів у 2019 році до 162,7 мільярда доларів у 2021 році. Це зростання було зумовлено переважно зростанням доходів від реклами та Google Cloud. Валовий прибуток також зросла відповідно до зростання доходів швидше, ніж витрати на дохід.

Операційний прибуток коливався протягом років. Витрати на компенсацію акцій та придбання сприяли зростанню операційних витрат, хоча були отримані прибутки від амортизації та амортизації.

Чистий прибуток збільшився з 34,3 мільярда доларів у 2019 році до 41,6 мільярда доларів у 2021 році. Потік грошей від операцій також стабільно зростав протягом 3 років.

Бізнес компанії сильно концентрується в Сполучених Штатах, хоча доходи з міжнародних ринків, особливо з Азії та ЄМА, зростають.

В цілому фінансові тенденції свідчать про постійне зростання та прибутковість Google протягом останніх 3 років.

Додаток В. Аналіз фінансового звіту Apple Inc

Сторінка фінансового звіту:

Apple Inc.				
CONDENSED CONSOLIDATED STATEMENTS OF OPERATIONS (Unaudited)				
(In millions, except number of shares which are reflected in thousands and per share amounts)				
	Three Months Ended		Six Months Ended	
	April 1, 2023	March 26, 2022	April 1, 2023	March 26, 2022
Net sales:				
Products	\$ 73,929	\$ 77,457	\$ 170,317	\$ 181,886
Services	20,907	19,821	41,673	39,337
Total net sales ⁽¹⁾	94,836	97,278	211,990	221,223
Cost of sales:				
Products	46,795	49,290	107,560	113,599
Services	6,065	5,429	12,122	10,822
Total cost of sales	52,860	54,719	119,682	124,421
Gross margin	41,976	42,559	92,308	96,802
Operating expenses:				
Research and development	7,457	6,387	15,166	12,693
Selling, general and administrative	6,201	6,193	12,808	12,642
Total operating expenses	13,658	12,580	27,974	25,335
Operating income	28,318	29,979	64,334	71,467
Other income/(expense), net	64	160	(329)	(87)
Income before provision for income taxes	28,382	30,139	64,005	71,380
Provision for income taxes	4,222	5,129	9,847	11,740
Net income	\$ 24,160	\$ 25,010	\$ 54,158	\$ 59,640
Earnings per share:				
Basic	\$ 1.53	\$ 1.54	\$ 3.42	\$ 3.65
Diluted	\$ 1.52	\$ 1.52	\$ 3.41	\$ 3.62
Shares used in computing earnings per share:				
Basic	15,787,154	16,278,802	15,839,939	16,335,263
Diluted	15,847,050	16,403,316	15,901,384	16,461,304
⁽¹⁾ Net sales by reportable segment:				
Americas	\$ 37,784	\$ 40,882	\$ 87,062	\$ 92,378
Europe	23,945	23,287	51,626	53,036
Greater China	17,812	18,343	41,717	44,126
Japan	7,176	7,724	13,931	14,831
Rest of Asia Pacific	8,119	7,042	17,654	16,852
Total net sales	\$ 94,836	\$ 97,278	\$ 211,990	\$ 221,223
⁽¹⁾ Net sales by category:				
iPhone	\$ 51,334	\$ 50,570	\$ 117,109	\$ 122,198
Mac	7,168	10,435	14,903	21,287
iPad	6,670	7,646	16,066	14,894
Wearables, Home and Accessories	8,757	8,806	22,239	23,507
Services	20,907	19,821	41,673	39,337
Total net sales	\$ 94,836	\$ 97,278	\$ 211,990	\$ 221,223

Результат:

Короткий фінансовий звіт компанії Apple Inc. базується на їх скорочених узагальнених зведених звітах про операції, баланси та грошові потоки за вказаний період. Ось деякі ключові моменти:

- Чистий обсяг продажів за три місяці, що закінчилися 1 квітня 2023 року, становив 94,8 мільярда доларів, порівняно з 97,3 мільярда доларів за той самий період 2022 року.
- Валовий прибуток за шість місяців, що закінчилися 1 квітня 2023 року, становив 92,3 мільярда доларів, порівняно з 96,8 мільярда доларів за той самий період 2022 року.
- Операційний прибуток за шість місяців, що закінчилися 1 квітня 2023 року, становив 64,3 мільярда доларів, порівняно з 71,5 мільярда доларів за той самий період 2022 року.
- Чистий прибуток за шість місяців, що закінчилися 1 квітня 2023 року, становив 54,2 мільярда доларів, порівняно з 59,6 мільярда доларів за той самий період 2022 року.
- **Скорочений зведений баланс:**
- Загальні поточні активи станом на 1 квітня 2023 року склали 112,9 мільярда доларів, порівняно з 135,4 мільярда доларів станом на 24 вересня 2022 року.
- Загальні необоротні активи станом на 1 квітня 2023 року склали 219,2 мільярда доларів, порівняно з 217,4 мільярда доларів станом на 24 вересня 2022 року.
- Загальні зобов'язання станом на 1 квітня 2023 року склали 270 мільярдів доларів, порівняно з 302,1 мільярда доларів станом на 24 вересня 2022 року.
- Загальний капітал акціонерів станом на 1 квітня 2023 року склав 62,2 мільярда доларів, порівняно з 50,7 мільярда доларів станом на 24 вересня 2022 року.
- **Скорочений зведений звіт про грошові потоки:**
- Гроші, згенеровані операційною діяльністю за шість місяців, що закінчилися 1 квітня 2023 року, становили 62,6 мільярда доларів, порівняно з 75,1 мільярда доларів за той самий період 2022 року.
- Гроші, згенеровані/використані в інвестиційній діяльності за шість місяців, що закінчилися 1 квітня 2023 року, становили 874 мільйони доларів, порівняно з втратою в розмірі 25,4 мільярда доларів за той самий період 2022 року.
- Гроші, використані в фінансовій діяльності за шість місяців, що закінчилися 1 квітня 2023 року, становили 61,3 мільярда доларів, порівняно з 56,5 мільярда доларів за той самий період 2022 року.

Важливо зауважити, що ці цифри виражені в мільйонах, за винятком кількості акцій та сум на акцію, які виражені в тисячах та на акцію, відповідно.