ТОВ «ТЕХНІЧНИЙ УНІВЕРСИТЕТ «МЕТІНВЕСТ ПОЛІТЕХНІКА»
Факультет автоматизації виробництва та цифрових технологій
Кафедра цифрових технологій та проєктно-аналітичних рішень

«Допущено до захисту»
Гарант ОПП
Павло САГАЙДА

КВАЛІФІКАЦІЙНА РОБОТА

на здобуття освітнього ступеня магістра

за підсумками виконання
освітньо-професійної програми
«Комп'ютерні науки та цифровий інтелект»
за спеціальністю 122 Комп`ютерні науки

на тему «Дослідження моделей, методів та інформаційних технологій використання агентів штучного інтелекту для автоматизації обробки даних»

Керівник роботи
Ірина ГЕТЬМАН

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

Здобувач
Максим КАРПЕНКО

Підсумкова оцінка за атестацію	

Голова ЕК
Антон КУДРЯВЦЕВ

ЗАПОРІЖЖЯ 2026

metinvest polytechnic

ТОВ «ТЕХНІЧНИЙ УНІВЕРСИТЕТ «МЕТІНВЕСТ ПОЛІТЕХНІКА»

Факультет	автоматизації виробництва та цифрових технологій
Кафедра	цифрових технологій та проектно-аналітичних рішень
Ступінь вищої освіти	магістр
Спеціальність	122 Комп'ютерні науки
ОПП	Комп'ютерні науки та цифровий інтелект

ЗАТВЕРДЖУЮ
Гарант ОПП
Павло САГАЙДА

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ МАГІСТРА

Карпенко Максиму Олександровичу
(прізвище, ім'я, по батькові здобувача)

1. Тема роботи *Дослідження моделей, методів та інформаційних технологій використання агентів штучного інтелекту для автоматизації обробки даних*
керівник роботи *Гетьман Ірина Анатоліївна, доцент, канд. техн. наук,*
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)
затверджені наказом Університету від 10.09.2025 р. №239/10.09.2025
2. Термін подання роботи 20.01.2026 р.
3. Вихідні дані до роботи *Навчальна література, державні стандарти, методична література з спеціальних дисциплін та підготовки кваліфікаційної роботи, науково-дослідницькі роботи з тематики автоматизації обробки й аналізу даних та методів цифрового інтелекту, літературні джерела, результати власних експериментів та досліджень, технологічні інструкції тощо*
4. Зміст пояснювальної записки (перелік питань) виконати *аналіз стану питання та сучасних підходів до автоматизації обробки вхідних електронних листів; розробити моделі та методи автоматизації обробки вхідних електронних листів з використанням агентів шучного інтелекту; розробити програмне забеспечення комплексу AI Email Parser Agent; провести експериментальні дослідження і аналіз результатів роботи програмного комплексу автоматизації обробки вхідних електронних повідомлень; виконати економічні розрахунки та оцінити ефективності впровадження розроблених рішень.*
5. Перелік графічного (демонстраційного) матеріалу (з точним зазначенням обов'язкових креслень): *мета і завдання дослідження; математичні моделі, методика дослідження; діаграми проекту програмної системи в нотації UML (діаграми прецедентів, класів, послідовностей, діяльності, потоків даних); результати експериментальних досліджень.*

6. Консультанти по роботі, із зазначенням розділів роботи, що стосуються їх.

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
Фаховий консультант	І.А. Гетьман, доц., канд. техн. наук		
Нормоконтроль	О.С. Касьянюк, ст. викл.		
Економічна частина	І.А. Гетьман, доц., канд. техн. наук		

7. Дата видачі завдання *10.11.2025*

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів роботи	Термін виконання етапів роботи
1	Розділ 1. Аналіз стану питання та сучасних підходів до автоматизації обробки вхідних електронних листів	20.01.2026 – 21.01.2026
2	Розділ 2. Розробка моделей та методів автоматизації обробки вхідних електронних листів з використанням агентів штучного інтелекту	22.01.2026 – 23.01.2026
3	Розділ 3. Розробка програмного комплексу AI Email Parser Agent	24.01.2026 – 25.01.2026
4	Розділ 4. Проведення та результати теоретичних та експериментальних досліджень програмного комплексу автоматизованої обробки вхідних електронних повідомлень	26.01.2026 – 27.01.2026
5	Розділ 5. Економічні розрахунки	27.01.2026
7	Оформлення текстової, програмної та графічної документації	28.01.2026
8	Оформлення записки. Підготовка докладу та презентації. Отримання відгуку та рецензії	28.01.2026
9	Захист проєкту	29.01.2026

Здобувач вищої освіти М.О. Карпенко
(ініціали та прізвище)

Керівник роботи, доц., канд. техн. наук _____ І.А. Гетьман
(підпис) (ініціали та прізвище)

РЕФЕРАТ

Карпенко М.О. Дослідження моделей, методів та інформаційних технологій використання агентів штучного інтелекту для автоматизації обробки даних.

Кваліфікаційна робота на здобуття освітнього ступеня магістра комп'ютерних наук за спеціальністю 122 Комп'ютерні науки. – ТОВ «ТЕХНІЧНИЙ УНІВЕРСИТЕТ «МЕТІНВЕСТ ПОЛІТЕХНІКА» МОН України, м. Запоріжжя, 2026.

Метою роботи є підвищення ефективності автоматизованої обробки вхідних електронних повідомлень у корпоративних інформаційних системах шляхом розробки інтелектуального агента для класифікації повідомлень та витягу структурованих даних. Об'єктом дослідження є процес обробки неструктурованої текстової інформації в CRM-системах. Предметом дослідження є моделі класифікації намірів повідомлень, методи витягу іменованих сутностей та інформаційні технології побудови інтелектуальних агентів. Методологія дослідження включає методи машинного навчання, обробки природної мови, агентно-орієнтоване проєктування та експериментальні дослідження.

У роботі проаналізовано сучасні підходи до автоматизації обробки текстових даних, розроблено архітектуру інтелектуального агента та програмний прототип для аналізу електронних листів. Реалізовано REST API та інтеграційний сценарій у середовищі SAP Cloud Platform Integration. Проведені експериментальні дослідження підтвердили ефективність запропонованих моделей і методів.

Результати роботи можуть бути використані для автоматизації процесу формування комерційних можливостей у CRM-системах. Подальший розвиток рішення передбачає розширення функціональності агента та донавчання моделей на реальних даних.

КЛЮЧОВІ СЛОВА: ІНТЕЛЕКТУАЛЬНИЙ АГЕНТ, ШТУЧНИЙ ІНТЕЛЕКТ, ОБРОБКА ПРИРОДНОЇ МОВИ, МАШИННЕ НАВЧАННЯ, КЛАСИФІКАЦІЯ ТЕКСТІВ, ВИТЯГ СУТНОСТЕЙ, CRM-СИСТЕМИ, SAP C4C

ABSTRACT

Karpenko M.O. Research on models, methods, and information technologies for using artificial intelligence agents to automate data processing.

Qualification work for obtaining the degree of Master of Computer Science in the specialty 122 Computer Science. – LLC «TECHNICAL UNIVERSITY «METINVEST POLYTECHNIC» MES of Ukraine, Zaporizhzhia, 2026.

The aim of the work is to improve the efficiency of automated processing of incoming electronic messages in corporate information systems by developing an intelligent agent for message classification and structured data extraction. The object of the research is the process of processing unstructured textual information in CRM systems. The subject of the research includes intent classification models, named entity recognition methods, and information technologies for building intelligent agents. The research methodology involves machine learning, natural language processing, agent-oriented design, and experimental evaluation.

The work analyzes modern approaches to text processing automation, proposes an intelligent agent architecture, and implements a software prototype for email analysis. A REST API and an integration scenario in the SAP Cloud Platform Integration environment have been developed. Experimental results confirm the effectiveness of the proposed solution.

The results can be applied to automate the creation of business opportunities in CRM systems. Further development includes extending the agent’s functionality and retraining models using real-world data.

KEYWORDS: INTELLIGENT AGENT, ARTIFICIAL INTELLIGENCE, NATURAL LANGUAGE PROCESSING, MACHINE LEARNING, TEXT CLASSIFICATION, NAMED ENTITY RECOGNITION, CRM SYSTEMS, SAP C4C

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

CRM – Customer Relationship Management (система управління взаємовідносинами з клієнтами);

SAP C4C – SAP Cloud for Customer (хмарна CRM-система SAP);

SAP CPI – SAP Cloud Platform Integration (інтеграційна платформа SAP);

ML – Machine Learning (машинне навчання);

NLP – Natural Language Processing (обробка природної мови);

NER – Named Entity Recognition (витяг іменованих сутностей);

API – Application Programming Interface (прикладний програмний інтерфейс);

ТЗ – технічне завдання;

ПЗ – програмне забезпечення;

IDE – Integrated Development Environment (інтегроване середовище розробки);

PoC – Proof of Concept (прототип для перевірки концепції);

EOL - End Of Life (кінець життєвого циклу продукту);

TF-IDF – Term Frequency–Inverse Document Frequency (метод векторного подання тексту);

ASGI – Asynchronous Server Gateway Interface (асинхронний інтерфейс серверних застосунків).

ЗМІСТ

ВСТУП 8
1 АНАЛІЗ СТАНУ ПИТАННЯ ТА СУЧАСНИХ ПІДХОДІВ ДО АВТОМАТИЗАЦІЇ ОБРОБКИ ВХІДНИХ ЕЛЕКТРОННИХ ЛИСТІВ..... 12
1.1 Аналіз процесу обробки неструктурованих даних у корпоративних системах управління взаємовідносинами з клієнтами (CRM) 12
1.2 Огляд існуючих методів та підходів до автоматизації обробки текстової інформації 18
1.2.1 Методи на основі правил (Rule-based) та регулярних виразів 20
1.2.2 Статистичні методи машинного навчання та NLP 22
1.2.3 Сучасні підходи на основі інтелектуальних агентів та LLM 25
1.3 Аналіз інструментальних засобів для реалізації системи 27
1.3.1 Інтеграційна платформа SAP Cloud Platform Integration (CPI) 28
1.3.2 Засоби розробки AI-компонентів (Python, FastAPI, spaCy) 29
1.4 Постановка задачі дослідження та формулювання вимог до системи 30
Висновки до розділу 1 31
2 РОЗРОБКА МОДЕЛЕЙ ТА МЕТОДІВ АВТОМАТИЗАЦІЇ ОБРОБКИ ВХІДНИХ ЕЛЕКТРОННИХ ЛИСТІВ З ВИКОРИСТАННЯМ АГЕНТІВ ШТУЧНОГО ІНТЕЛЕКТУ 33
2.1 Формалізація задачі автоматизованої обробки вхідного потоку повідомлень 33
2.2 Математична модель класифікації намірів (Intent Classification) на основі логістичної регресії 36
2.3 Метод вилучення іменованих сутностей (NER) та модель нормалізації атрибутів 42

2.4 Алгоритм функціонування мультиагентної системи обробки даних 43
2.5 Вибір критеріїв якості та метрик оцінювання ефективності моделей 47
Висновки до розділу 2 49
3 РОЗРОБКА ПРОГРАМНОГО КОМПЛЕКСУ AI EMAIL PARSER AGENT 51
3.1 Розробка бізнес-процесів автоматизації створення об’єктів CRM на основі інтелектуальної обробки вхідних електронних повідомлень 51
3.2 Розробка логічної та фізичної моделі програмного комплексу для автоматизації обробки вхідних електронних повідомлень у CRM-системі 71
3.3 Програмна реалізація інтелектуального агента 91
3.3.1 Структура проєкту та опис основних модулів 92
3.3.2 Реалізація API для аналізу тексту (FastAPI) 94
3.3.3 Технологічний стек програмного комплексу 95
3.4. Реалізація інтеграційного сценарію (iFlow) в середовищі SAP CPI 97
Висновки до розділу 3 99
4 ПРОВЕДЕННЯ ТА РЕЗУЛЬТАТИ ТЕОРЕТИЧНИХ ТА ЕКСПЕРИМЕНТАЛЬНИХ ДОСЛІДЖЕНЬ ПРОГРАМНОГО КОМПЛЕКСУ АВТОМАТИЗОВАНОЇ ОБРОБКИ ВХІДНИХ ЕЛЕКТРОННИХ ПОВІДОМЛЕНЬ 101
4.1 Мета, завдання та загальна організація експериментальних досліджень 101
4.2 Опис програмного комплексу та середовища проведення експериментів 102

4.3 Дослідження процесу класифікації релевантності електронних повідомлень 105
4.3.1 Дослідження процесу класифікації релевантності електронних повідомлень 109
4.4 Дослідження якості витягу структурованих даних з електронних листів 112
4.5 Аналіз результатів експериментальних досліджень 116
4.6 Практичні рекомендації щодо використання програмного комплексу 119
Висновки до розділу 4 121
5. ЕКОНОМІЧНЕ ОБҐРУНТУВАННЯ ВПРОВАДЖЕННЯ ПРОГРАМНОГО КОМПЛЕКСУ AI EMAIL PARSER AGENT 123
5.1 Загальна характеристика економічного ефекту від впровадження програмного комплексу 123
5.2 Оцінка трудових витрат до та після впровадження системи. 124
5.3 Орієнтовний економічний ефект від автоматизації обробки електронних повідомлень 125
5.4 Витрати на впровадження та оцінка доцільності використання рішення 128
Висновки до розділу 5 129
ВИСНОВКИ 131
ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ 134
ДОДАТОК А 137
ДОДАТОК Б 140

ВСТУП

Актуальність теми.

У сучасних умовах цифрової трансформації великих промислових підприємств, зокрема підприємств групи «Метінвест», обсяги вхідної електронної кореспонденції від контрагентів та потенційних клієнтів мають стійку тенденцію до зростання. Значна частина таких повідомлень містить комерційні запити, що потребують оперативної обробки з метою формування комерційних можливостей (Opportunity) у корпоративних CRM-системах, зокрема SAP Cloud for Customer (SAP C4C). На практиці даний процес переважно виконується вручну менеджерами відділів продажів, що призводить до додаткових часових витрат, підвищення навантаження на персонал та ризику помилок, зумовлених людським фактором, а також до затримок у реагуванні на запити ринку [17].

Існуючі засоби автоматизації обробки електронної кореспонденції, які базуються на жорстко формалізованих правилах і шаблонах, демонструють обмежену ефективність у роботі з неструктурованими текстовими даними природною мовою. Така ситуація зумовлює наявність суперечності між потребою бізнесу в гнучкій та масштабованій автоматизації і можливостями традиційних підходів. Розв’язання цієї проблеми потребує дослідження моделей і методів використання агентів штучного інтелекту, здатних автономно інтерпретувати зміст вхідних повідомлень, класифікувати їх за релевантністю та здійснювати витяг ключових сутностей, необхідних для підтримки бізнес-процесів [1, 3]. У зв’язку з цим розробка інтелектуального агента для автоматизації обробки вхідних

електронних листів є актуальним науково-практичним завданням у контексті розвитку сучасних корпоративних інформаційних систем.

Зв’язок роботи з науково-технічними програмами, планами, темами.

Обраний напрям дослідження пов’язаний із програмами цифровізації підприємств групи «Метінвест», спрямованими на впровадження інтелектуальних сервісів автоматизації бізнес-процесів, зокрема рішень класу RPA та AI-агентів. Робота також відповідає науковим напрямам кафедри цифрових технологій та проєктно-аналітичних рішень ТУ «Метінвест Політехніка», що орієнтовані на розвиток моделей аналізу великих даних та методів машинного навчання в корпоративних інформаційних системах.

Мета і завдання дослідження.

Метою роботи є підвищення ефективності бізнес-процесу формування комерційних можливостей у CRM-системі шляхом дослідження та розробки моделей, методів і інформаційної технології функціонування інтелектуального агента для автоматизованої обробки вхідних електронних повідомлень.

Об’єкт дослідження - процес автоматизованої обробки вхідної неструктурованої текстової інформації в корпоративних інформаційних системах.

Предмет дослідження - моделі класифікації намірів повідомлень, методи витягу сутностей (Named Entity Recognition) та архітектурні рішення побудови інтелектуальних агентів для формування структурованих даних.

Для досягнення поставленої мети в роботі передбачено розв’язання таких завдань:

1. Проаналізувати сучасні підходи до побудови мультиагентних систем і методи обробки природної мови.

2. Розробити модель класифікації релевантності електронних листів на основі методів машинного навчання та векторного представлення текстів.

3. Розробити модель витягу структурованих даних з неструктурованого тексту на основі гібридного підходу, що поєднує нейромережеві методи та правила.

4. Спроєктувати архітектурну модель та реалізувати програмний прототип інтелектуального агента з використанням технологій Python та FastAPI.

5. Провести експериментальне дослідження ефективності розробленого рішення на підготовленому датасеті.

Методи дослідження.

У роботі використано методи теорії штучного інтелекту та мультиагентних систем для проєктування архітектури інтелектуального агента; методи машинного навчання, зокрема логістичну регресію та TF-IDF-векторизацію, для класифікації релевантності повідомлень; методи обробки природної мови та регулярні вирази - для витягу ключових сутностей із текстових даних.

Наукова новизна роботи полягає в удосконаленні архітектурної моделі інтелектуального агента-оркестратора, яка реалізує ієрархічний цикл обробки повідомлення (сприйняття - аналіз - прийняття рішення) з урахуванням специфіки корпоративних текстових даних металургійної галузі, а також у подальшому розвитку моделей витягу сутностей шляхом застосування гібридного підходу, що підвищує точність ідентифікації товарних позицій у текстах природною мовою.

Практичне значення отриманих результатів.

Практична значущість роботи полягає у створенні програмного прототипу (Proof of Concept) інтелектуального агента «AI Email Parser

Agent for SAP C4C Opportunity», призначеного для валідації моделей класифікації вхідних електронних листів і витягу структурованих даних у корпоративних інформаційних системах. Розроблений прототип формує стандартизований машинозчитуваний результат у форматі JSON, придатний для подальшої інтеграції з CRM-системами через програмні інтерфейси. У межах дослідження апробовано методику автоматизованої обробки вибірки з 600 повідомлень і досягнуто точності класифікації на рівні 95%, що підтверджує ефективність обраних підходів. Отримані результати створюють експериментальну та архітектурну основу для подальшої промислової реалізації end-to-end сценарію автоматизації з використанням корпоративного інтеграційного шару.

Особистий внесок магістранта.

Робота виконана автором самостійно. Особистий внесок полягає у розробці алгоритмічної логіки інтелектуального агента, генерації та підготовці датасету, навчанні моделей машинного навчання та програмній реалізації прототипу. Використання сторонніх бібліотек супроводжується відповідними посиланнями. Робота пройшла перевірку на відсутність академічного плагіату.

Апробація результатів.

Основні результати дослідження доповідались та обговорювались на наукових семінарах кафедри цифрових технологій та проєктно-аналітичних рішень ТУ «Метінвест Політехніка» у 2025 році, а також на конференціях.

1 АНАЛІЗ СТАНУ ПИТАННЯ ТА СУЧАСНИХ ПІДХОДІВ ДО АВТОМАТИЗАЦІЇ ОБРОБКИ ВХІДНИХ ЕЛЕКТРОННИХ ЛИСТІВ

1.1 Аналіз процесу обробки неструктурованих даних у корпоративних системах управління взаємовідносинами з клієнтами (CRM)

У сучасних корпоративних інформаційних системах управління взаємовідносинами з клієнтами (Customer Relationship Management, CRM) обробка вхідних інформаційних потоків є критично важливою складовою підтримки бізнес-процесів продажу, сервісного обслуговування та управлінської аналітики. CRM-системи забезпечують централізоване зберігання та актуалізацію даних про клієнтів, історію взаємодії з ними, комерційні можливості, етапи угод та супровідні бізнес-процеси, що дозволяє формувати цілісне уявлення про стан взаємовідносин з контрагентами. Разом з тим ефективність таких систем безпосередньо залежить від якості, повноти та своєчасності первинних даних, які надходять із зовнішніх джерел. У реальних корпоративних умовах значна частина цієї інформації має неструктурований або слабкоструктурований характер, що проявляється у вигляді текстових повідомлень природною мовою, зокрема електронної кореспонденції, службових повідомлень та супровідних документів. Як зазначається у стандартах з управління інформаційними системами та програмною інженерією, робота з неструктурованими даними суттєво ускладнює автоматизацію процесів і потребує застосування спеціалізованих методів аналізу та інтерпретації інформації [18], [35]. Відсутність формалізованої структури даних призводить до зростання

навантаження на користувачів CRM-систем, збільшення часу обробки запитів клієнтів і підвищення ризику помилок, що негативно впливає на ефективність бізнес-процесів і якість управлінських рішень.

У структурі вхідних інформаційних потоків CRM-систем особливе місце займає електронна кореспонденція, яка в більшості випадків виступає первинним каналом комунікації між клієнтом і компанією. Саме через електронні листи клієнти надсилають запити на розрахунок вартості продукції, уточнюють умови співпраці, передають технічні вимоги або ініціюють початок переговорів щодо потенційних угод. На відміну від формалізованих каналів взаємодії, таких як веб-форми або інтеграційні повідомлення між інформаційними системами, електронні листи створюються у довільній формі природною мовою та не підпорядковуються єдиним стандартам структурування. Це призводить до значної варіативності змісту повідомлень, використання різних стилів викладення, скорочень, неформальних формулювань і навіть орфографічних помилок, що ускладнює автоматичне виділення бізнес-значущої інформації. Дослідження в галузі обробки природної мови підкреслюють, що саме така різноманітність лінгвістичних конструкцій є однією з основних причин низької ефективності простих алгоритмів аналізу текстів у реальних бізнес-сценаріях [3], [7].

З точки зору інформаційних технологій електронні листи належать до класу неструктурованих або слабкоструктурованих даних, оскільки їх зміст не має фіксованої схеми представлення та може поєднувати різні типи інформації в одному повідомленні. В одному листі клієнт може одночасно вказувати назву компанії, контактну особу, перелік товарних позицій, кількісні параметри, одиниці виміру, а також додаткові коментарі або умови співпраці, причому ці елементи можуть бути розташовані у довільному порядку

або взагалі неявно присутні у тексті. Така особливість ускладнює пряме відображення вмісту листа у структуровані атрибути CRM-системи без попереднього аналізу та інтерпретації тексту. У працях, присвячених аналізу текстових даних і інформаційному пошуку, зазначається, що варіативність лексичних, синтаксичних і семантичних форм є ключовим фактором, який обмежує застосування традиційних методів обробки даних і потребує використання спеціалізованих підходів до аналізу природної мови [4], [6].

У практиці використання CRM-систем обробка вхідної електронної кореспонденції здебільшого здійснюється вручну або з мінімальним рівнем автоматизації, що підтверджується як галузевими аналітичними звітами, так і дослідженнями з впровадження штучного інтелекту в бізнес-процесах [29], [31]. Узагальнену схему такого процесу обробки вхідних електронних повідомлень у CRM-системі без використання автоматизованих засобів наведено на рисунку 1.1. Типовий сценарій передбачає, що співробітник відділу продажів або служби підтримки відкриває електронний лист, аналізує його зміст, інтерпретує комерційний намір клієнта та вручну переносить релевантну інформацію до CRM-системи, створюючи або оновлюючи об’єкти типу Opportunity. При цьому процес інтерпретації тексту значною мірою залежить від досвіду та суб’єктивного сприйняття конкретного користувача, що суперечить принципам стандартизації та відтворюваності бізнес-процесів, які декларуються у підходах програмної інженерії та управління якістю програмних систем [34], [35]. Як зазначається у роботах з аналізу корпоративних інформаційних систем, така залежність від людського фактору призводить до непослідовності даних, накопичення помилок та

зниження достовірності аналітичної інформації, що формується на основі даних CRM [18], [29].

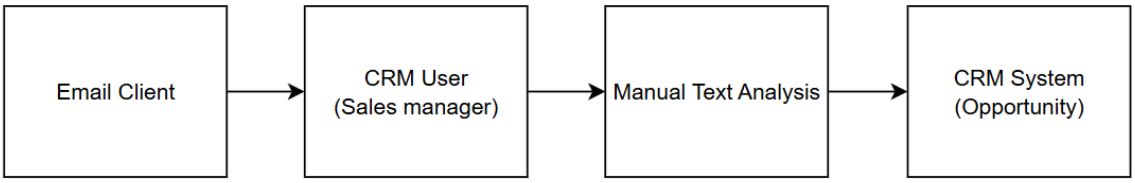

Рисунок 1.1 – Узагальнена схема процесу обробки вхідної електронної кореспонденції в CRM-системі без автоматизації.

Ще однією суттєвою проблемою ручної та напівавтоматичної обробки електронних листів у CRM-системах є відсутність єдиного формалізованого підходу до структурування отриманої з тексту інформації, що прямо впливає на якість даних у системі. Навіть за наявності внутрішніх регламентів різні підрозділи або окремі співробітники можуть по-різному інтерпретувати зміст одного й того самого повідомлення, виділяючи різні атрибути як бізнес-значущі, що не відповідає принципам узгодженості та цілісності даних, описаним у стандартах інформаційних технологій [18]. У результаті в CRM накопичуються записи з різним рівнем деталізації, неповними або суперечливими значеннями полів, що ускладнює подальший аналіз продажів, прогнозування та побудову звітності. Як підкреслюється у працях, присвячених застосуванню аналітики та машинного навчання в корпоративних системах, якість вихідних управлінських рішень безпосередньо залежить від якості вхідних даних, а неструктуровані та неуніфіковані джерела інформації є однією з головних причин деградації аналітичних моделей [6], [10]. У контексті CRM це означає, що навіть наявність розвинених інструментів аналітики не дає очікуваного ефекту без попередньої інтелектуальної обробки та

уніфікації текстових даних, отриманих з електронної кореспонденції [29].

З технічної точки зору більшість сучасних CRM-систем спроєктовані для роботи переважно зі структурованими даними, які зберігаються у реляційних базах даних та передаються між компонентами системи через формалізовані програмні інтерфейси, що відповідає класичним підходам до побудови корпоративних інформаційних систем [34], [35]. Такі платформи ефективно оперують числовими показниками, довідниками, статусами та заздалегідь визначеними атрибутами бізнес-об’єктів, однак не мають вбудованих механізмів глибокого семантичного аналізу текстів природною мовою. Як зазначається у роботах з інформаційного пошуку та обробки текстів, пряме застосування традиційних методів роботи з даними до неструктурованого тексту призводить до втрати значної частини змістовної інформації або потребує значних зусиль з попередньої підготовки даних [6], [4]. У результаті виникає розрив між реальними бізнес-комунікаціями, які відбуваються у формі текстових повідомлень, та можливостями стандартних CRM-платформ щодо їх автоматизованої інтерпретації і використання в бізнес-процесах, що додатково підсилює необхідність застосування спеціалізованих інтелектуальних методів аналізу текстових даних [29].

Для часткового подолання зазначених обмежень у практиці корпоративних інформаційних систем інколи застосовуються прості механізми автоматизації, зокрема фільтрація електронних листів за відправником, темою повідомлення або наявністю певних ключових слів. Такі підходи дозволяють зменшити обсяг ручної роботи на початковому етапі обробки кореспонденції, проте не вирішують проблему глибокого аналізу змісту повідомлень. Як показано у класичних роботах з автоматичної категоризації текстів і машинного

навчання для обробки текстової інформації, правило-орієнтовані методи мають низьку стійкість до варіативності мови, синонімії та зміни контексту, що є характерним для реальних бізнес-комунікацій [5], [20]. Крім того, зі зростанням обсягу вхідних повідомлень і появою багатомовних сценаріїв підтримка таких правил стає складною та ресурсоємною, що суперечить вимогам масштабованості й адаптивності, які висуваються до сучасних CRM-рішень [29], [31].

Узагальнюючи наведене, можна зробити висновок, що існуючі підходи до обробки неструктурованих текстових даних у CRM-системах не відповідають вимогам сучасних корпоративних бізнес-процесів, які орієнтовані на швидкість реагування, узгодженість даних та масштабованість рішень. Ручна та правило-орієнтована обробка електронної кореспонденції не забезпечує необхідного рівня точності, відтворюваності та адаптивності, особливо в умовах зростання обсягів вхідних повідомлень і ускладнення клієнтських сценаріїв [29], [31]. Як підкреслюється у працях з аналізу текстових даних і штучного інтелекту, подолання зазначених обмежень можливе лише за рахунок застосування методів автоматизованого аналізу природної мови та інтелектуальних агентів, здатних інтерпретувати зміст тексту, виділяти бізнес-значущі сутності та інтегрувати результати аналізу безпосередньо у процеси корпоративних інформаційних систем [3], [4], [11]. Це обґрунтовує доцільність подальшого розгляду сучасних методів і підходів до автоматизації обробки текстової інформації, що є предметом наступного підрозділу.

1.2 Огляд існуючих методів та підходів до автоматизації обробки текстової інформації

Автоматизація обробки текстової інформації в корпоративних інформаційних системах є окремим науково-практичним напрямом, що активно розвивається на перетині машинного навчання, обробки природної мови та інженерії програмного забезпечення. У контексті CRM-систем така автоматизація спрямована на перетворення неструктурованих текстових повідомлень, зокрема електронних листів, у формалізовані дані, придатні для подальшої обробки бізнес-процесами підприємства. Як зазначають D. Jurafsky та J. Martin, будь-яка задача аналізу тексту починається з вибору відповідного підходу до подання та інтерпретації мовних даних, що безпосередньо впливає на точність класифікації, повноту витягу сутностей та загальну ефективність системи [3].

У наукових та прикладних джерелах прийнято виділяти декілька основних класів методів автоматизації текстової обробки. Найбільш ранні підходи базуються на використанні жорстко заданих правил та шаблонів, сформованих експертами предметної області. Такі rule-based системи виконують аналіз тексту шляхом пошуку ключових слів, регулярних виразів та лінгвістичних конструкцій, що відповідають заздалегідь визначеним сценаріям. Основною перевагою даного підходу є прозорість логіки прийняття рішень та відсутність необхідності у навчальних вибірках. Водночас його суттєвим недоліком є низька адаптивність до зміни структури повідомлень, мовної варіативності та наявності шуму в даних, що обмежує можливості практичного застосування в умовах реальних корпоративних потоків інформації [4].

Подальший розвиток автоматизації обробки текстових даних пов’язаний із впровадженням статистичних методів машинного навчання. У межах цього підходу текстові повідомлення представляються у вигляді числових векторів з використанням методів частотного аналізу, таких як Bag-of-Words або TF-IDF, після чого застосовуються класичні алгоритми класифікації, зокрема логістична регресія, метод опорних векторів або наївний баєсівський класифікатор. Як показано в роботах F. Sebastiani, статистичні методи демонструють достатньо високу ефективність у задачах автоматичної категоризації текстів за умови коректної побудови навчальної вибірки та належного налаштування моделей [5]. Разом з тим такі підходи часто потребують додаткових механізмів для витягу структурованих атрибутів і не забезпечують повної автономності аналізу складних текстових повідомлень.

Окрему групу становлять сучасні підходи, що ґрунтуються на використанні методів глибокого навчання та великих мовних моделей. Дані рішення дозволяють враховувати семантичні залежності між словами, контекстні зв’язки та мовні особливості, що істотно підвищує точність аналізу. Проте впровадження таких моделей у корпоративних інформаційних системах пов’язане з рядом практичних обмежень, зокрема високими обчислювальними витратами, складністю інтерпретації результатів та залежністю від зовнішніх сервісів або хмарної інфраструктури. У зв’язку з цим застосування великих мовних моделей у бізнес-критичних сценаріях потребує додаткового обґрунтування та адаптації [20].

Альтернативним напрямом розвитку є використання інтелектуальних агентів, які поєднують різні методи аналізу тексту в межах ієрархічної або модульної архітектури. У таких системах окремі агенти або модулі відповідають за класифікацію наміру повідомлення,

витяг іменованих сутностей, нормалізацію даних та прийняття рішень щодо подальших бізнес-дій. Агентний підхід дозволяє забезпечити автономність обробки, гнучкість архітектури та можливість поступового розширення функціональності системи без повної перебудови програмного комплексу. Саме такі рішення розглядаються як найбільш перспективні для інтеграції з корпоративними CRM-системами, де важливо поєднати точність аналізу з надійністю та керованістю процесів [4], [5].

Таким чином, аналіз існуючих методів і підходів до автоматизації обробки текстової інформації свідчить про відсутність універсального рішення, придатного для всіх сценаріїв використання. Rule-based системи забезпечують простоту та прозорість, статистичні методи машинного навчання - прийнятний баланс між точністю та обчислювальною складністю, тоді як інтелектуальні агентні системи дозволяють інтегрувати різні підходи в межах єдиної архітектури. Це зумовлює доцільність розробки гібридного програмного комплексу, орієнтованого на автоматизацію обробки електронних повідомлень у корпоративному середовищі, що і визначає напрям подальших досліджень у межах даної роботи.

1.2.1 Методи на основі правил (Rule-based) та регулярних виразів

Найбільш ранні та концептуально прості підходи до автоматизації обробки текстової інформації базуються на використанні явно заданих правил, шаблонів та регулярних виразів. У таких системах логіка аналізу тексту визначається наперед у вигляді

набору умов, які описують допустимі мовні конструкції, ключові слова або формальні шаблони даних. Подібні підходи широко застосовувалися на початкових етапах розвитку систем обробки природної мови, зокрема для задач фільтрації повідомлень, пошуку ключових фраз та вилучення простих атрибутів із тексту [6], [7]. Регулярні вирази дозволяють ефективно ідентифікувати формалізовані елементи, такі як числові значення, одиниці виміру або стандартні скорочення, що робить їх привабливими для первинної автоматизації бізнес-процесів у CRM-системах [4]. Водночас, як підкреслює F. Sebastiani, rule-based підходи практично не масштабуються при зростанні різноманітності текстових формулювань і вимагають постійного ручного супроводу, що істотно обмежує їх застосування у реальних корпоративних середовищах [5]. Типові результати застосування rule-based механізмів аналізу електронної кореспонденції, а також характерні помилки у вигляді пропусків намірів та хибних спрацьовувань, узагальнено проілюстровано на рисунку 1.2.

Попри простоту реалізації та низькі обчислювальні витрати, методи на основі правил мають суттєві концептуальні та практичні обмеження. Основною проблемою є їхня жорстка залежність від наперед визначених шаблонів, що робить такі системи чутливими до будь-яких відхилень у формулюваннях тексту. Навіть незначні зміни у порядку слів, використання синонімів або мовних скорочень можуть призводити до втрати коректного розпізнавання інформації, що підтверджується результатами досліджень у галузі автоматичної категоризації текстів [5], [20]. Крім того, rule-based системи практично не здатні враховувати контекст повідомлення, що є критично важливим для бізнес-комунікацій, де значення окремих фраз визначається загальним змістом листа, а не лише наявністю

ключових слів [3], [9]. У корпоративних CRM-системах це призводить до високої кількості хибнопозитивних та хибнонегативних спрацьовувань, а також до необхідності постійного розширення та підтримки правил вручну, що знижує економічну доцільність таких рішень у довгостроковій перспективі [29], [31].

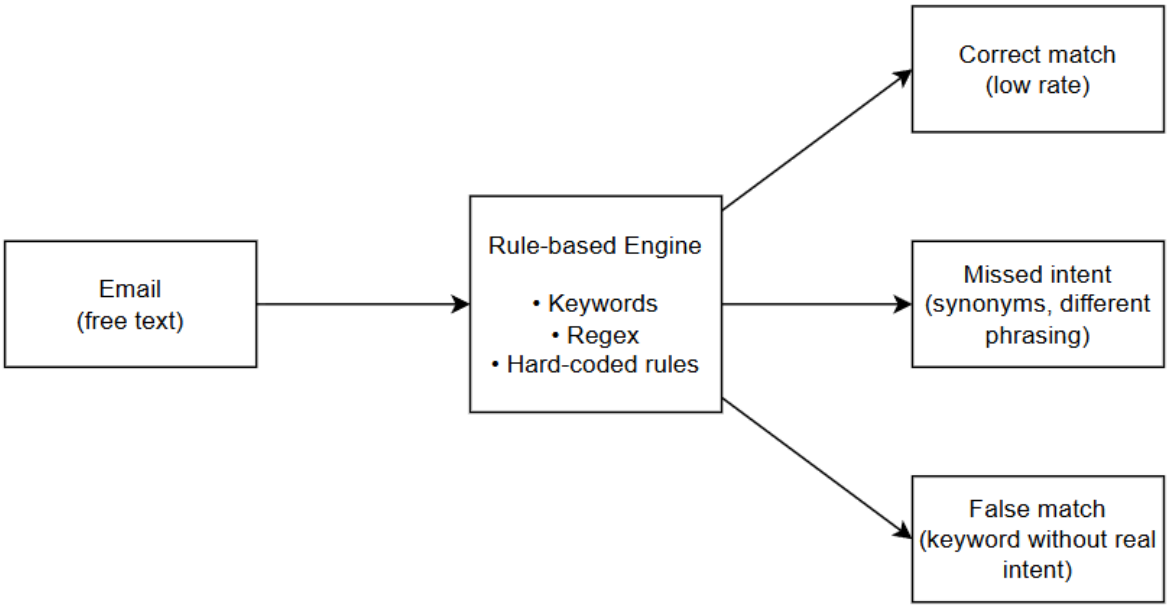

Рисунок 1.2 - Обмеження rule-based підходів при обробці електронної кореспонденції в CRM-системі

1.2.2 Статистичні методи машинного навчання та NLP

Подальший розвиток автоматизації обробки текстової інформації пов’язаний із застосуванням статистичних методів машинного навчання, які ґрунтуються на аналізі корпусів текстових даних та побудові узагальнених моделей замість жорстко заданих

правил. У таких підходах текстові повідомлення попередньо перетворюються у формалізоване подання у вигляді векторів ознак, що дозволяє застосовувати класичні алгоритми класифікації та аналізу даних. Однією з найбільш поширених моделей подання тексту є векторна просторова модель (Vector Space Model), у межах якої документи описуються через статистичні характеристики термінів, зокрема з використанням зважування TF-IDF, що детально описано в роботах C. Manning та співавторів [6], а також C. Aggarwal [4]. На основі такого подання можливе навчання моделей логістичної регресії, SVM або наївного баєсівського класифікатора для задач визначення наміру повідомлення або його тематичної категорії [5], [20]. Перевагою статистичних методів є здатність враховувати загальні мовні закономірності та адаптуватися до різних формулювань тексту, що робить їх більш придатними для застосування у CRM-системах порівняно з rule-based підходами, особливо за наявності достатнього обсягу навчальних даних [14], [21]. Типовий конвеєр статистичної обробки текстових повідомлень у задачах CRM, який включає етапи попередньої обробки, векторизації та машинного навчання, наведено на рисунку 1.3.

Разом із перевагами статистичні методи машинного навчання мають і низку принципових обмежень, які необхідно враховувати при їх використанні в корпоративних інформаційних системах. З одного боку, такі підходи демонструють значно кращу узагальнювальну здатність порівняно з rule-based рішеннями, оскільки модель навчається на прикладах і враховує частотні та контекстні закономірності мови, що підтверджується результатами класичних досліджень у сфері текстової класифікації [5], [20]. З іншого боку, якість роботи статистичних моделей безпосередньо залежить від обсягу та репрезентативності навчального корпусу, а також від

коректності попередньої обробки тексту, включаючи токенізацію, нормалізацію та видалення шуму [3], [7]. У бізнес-сценаріях, зокрема в CRM-системах, це створює додаткові труднощі, оскільки реальні електронні листи часто містять орфографічні помилки, неформальні скорочення та змішування мов, що негативно впливає на стабільність статистичних моделей [9], [23]. Крім того, класичні ML-підходи зазвичай працюють із фіксованим набором ознак і не забезпечують глибокого семантичного розуміння тексту, обмежуючись оцінкою статистичної подібності, що підкреслюється у працях Y. Goldberg та C. Bishop [8], [10].

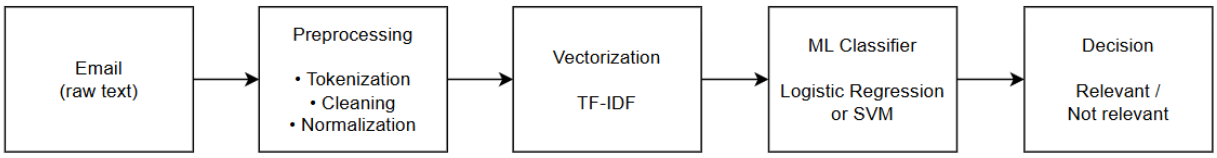

Рисунок 1.3 - Типовий конвеєр статистичної обробки тексту в задачах CRM

Додатково слід зазначити, що статистичні методи машинного навчання та класичні підходи NLP є компромісним рішенням між простотою rule-based систем і складністю сучасних нейромережевих моделей. З практичної точки зору вони добре піддаються інтерпретації, оскільки результати роботи таких моделей можуть бути пояснені через ваги ознак, внесок окремих термінів або значення ймовірностей класів, що особливо важливо для корпоративних систем прийняття рішень [10], [14]. Саме ця властивість робить статистичні моделі привабливими для використання в бізнес-середовищі, де необхідно забезпечити прозорість та контрольованість автоматизованих рішень відповідно до вимог

стандартів якості та безпеки програмного забезпечення [19], [33]. Водночас у наукових оглядах підкреслюється, що зі зростанням складності предметної області та кількості сценаріїв взаємодії ефективність таких підходів поступово знижується, оскільки вони не здатні повною мірою враховувати глибокі семантичні зв’язки та контекстуальні залежності в тексті [8], [9]. Це обумовлює необхідність подальшого розвитку систем обробки текстів у напрямі більш гнучких та інтелектуальних архітектур, що поєднують класичні ML-методи з агентними та нейромережевими підходами.

1.2.3 Сучасні підходи на основі інтелектуальних агентів та LLM

Сучасний етап розвитку автоматизації обробки текстової інформації характеризується переходом від окремих моделей машинного навчання до комплексних інтелектуальних систем, побудованих за агентним принципом. У таких системах обробка тексту розглядається не як одноразовий акт класифікації, а як послідовність взаємопов’язаних дій, що виконуються автономними програмними агентами відповідно до заданої мети. Як зазначається у працях M. Wooldridge та J. Ferber, інтелектуальний агент - це програмна сутність, здатна сприймати вхідні дані, приймати рішення та виконувати дії без прямого втручання користувача [11], [12]. У контексті CRM-систем такі агенти можуть автономно аналізувати вхідну електронну кореспонденцію, визначати її бізнес-релевантність, витягувати ключові атрибути та ініціювати подальші бізнес-процеси, що суттєво знижує навантаження на персонал і мінімізує вплив людського фактору [29], [30]. Узагальнену агентну схему

автоматизованої обробки електронних листів у CRM-системі, в якій інтелектуальний агент виконує функції класифікації наміру, витягу сутностей та прийняття рішень, наведено на рисунку 1.4.

Подальший розвиток агентних підходів тісно пов’язаний із появою великих мовних моделей (Large Language Models, LLM), які забезпечують якісно новий рівень аналізу текстів природною мовою. Такі моделі, побудовані на трансформерних архітектурах, здатні враховувати широкий контекст повідомлення, семантичні зв’язки між фрагментами тексту та неявні наміри автора, що докладно описано у роботах A. Vaswani та співавторів, а також J. Devlin і колег [27], [26]. На відміну від класичних статистичних моделей, LLM не потребують ручного проєктування ознак і можуть застосовуватися до різноманітних мовних сценаріїв, включаючи багатомовні та доменно-специфічні тексти, що є суттєвою перевагою для корпоративних CRM-систем [28], [30]. Водночас у наукових і прикладних дослідженнях підкреслюється, що використання LLM у промислових інформаційних системах пов’язане з низкою викликів, зокрема високими обчислювальними витратами, складністю контролю якості рішень та питаннями безпеки й конфіденційності даних, що є критичними для корпоративного середовища [29], [31], [33]. Саме тому на практиці дедалі частіше застосовуються гібридні підходи, у межах яких інтелектуальні агенти поєднують класичні ML-моделі, правила та LLM, забезпечуючи баланс між точністю, керованістю та продуктивністю системи.

У практичних корпоративних реалізаціях інтелектуальні агенти на базі LLM рідко використовуються як єдиний компонент системи, а зазвичай інтегруються в гібридні архітектури, де їм відводиться роль семантичного аналізу та підтримки прийняття рішень. Такий підхід дозволяє поєднати гнучкість і контекстну чутливість мовних моделей

із детермінованістю класичних алгоритмів і бізнес-правил, що є критично важливим для CRM-систем, орієнтованих на стабільну обробку великого потоку повідомлень [11], [29], [30]. Крім того, агентна організація системи спрощує масштабування та подальший розвиток програмного комплексу, оскільки окремі агенти можуть незалежно модифікуватися або замінюватися без порушення загальної логіки роботи рішення [12], [13].

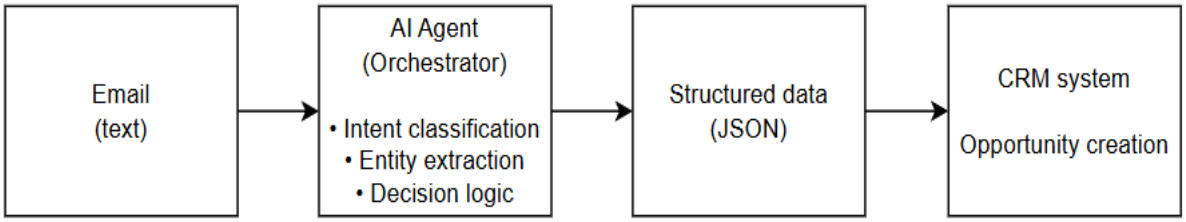

Рисунок 1.4 - Агентна схема автоматизованої обробки електронних листів у CRM-системі

1.3 Аналіз інструментальних засобів для реалізації системи

Реалізація системи автоматизованої обробки електронної кореспонденції в корпоративному CRM-середовищі вимагає використання інструментальних засобів, які забезпечують надійну інтеграцію між бізнес-системами, підтримку асинхронного обміну даними та можливість розгортання інтелектуальних компонентів поза межами ядра CRM. У сучасних корпоративних IT-ландшафтах такі вимоги обумовлюють необхідність застосування спеціалізованих інтеграційних платформ та гнучких технологій розробки, здатних працювати з REST-інтерфейсами, структурованими форматами

даних і сервісною архітектурою. Як зазначається у документації SAP та аналітичних оглядах з корпоративної автоматизації, відокремлення інтеграційної логіки від бізнес-логіки CRM дозволяє зменшити складність системи, підвищити її масштабованість та спростити подальший розвиток [17], [32], [35]. Саме тому доцільним є використання інтеграційної платформи як проміжної ланки між CRM-системою та зовнішнім інтелектуальним агентом, реалізованим із застосуванням сучасних засобів розробки AI-компонентів.

1.3.1 Інтеграційна платформа SAP Cloud Platform Integration (CPI)

Інтеграційна платформа SAP Cloud Platform Integration (CPI) призначена для організації обміну даними між корпоративними системами в межах сервісно-орієнтованої та подієвої архітектури. Вона забезпечує маршрутизацію повідомлень, трансформацію форматів даних та реалізацію інтеграційних сценаріїв без необхідності прямого втручання у бізнес-логіку CRM-системи [17], [32]. Використання SAP CPI як проміжного шару дозволяє ізолювати зовнішні інтелектуальні сервіси від внутрішньої структури SAP C4C, що відповідає рекомендаціям щодо побудови масштабованих та безпечних корпоративних IT-рішень [35].

1.3.2 Засоби розробки AI-компонентів (Python, FastAPI, spaCy)

Для реалізації інтелектуального компонента системи доцільним є використання мов та фреймворків, які забезпечують гнучкість розробки, розвинену екосистему бібліотек для обробки природної мови та можливість простої інтеграції з корпоративними платформами через стандартизовані інтерфейси. Мова програмування Python на сьогодні є де-факто стандартом у галузі машинного навчання та NLP, що обумовлено її лаконічним синтаксисом, широкою підтримкою наукових бібліотек та наявністю великої кількості перевірених рішень для задач класифікації текстів і вилучення сутностей [7], [14], [34]. Для організації взаємодії інтелектуального агента з інтеграційною платформою SAP CPI доцільно використовувати вебфреймворк FastAPI, який дозволяє швидко створювати REST-сервіси з чітко формалізованими контрактами обміну даними, автоматичною валідацією вхідних повідомлень та високою продуктивністю обробки запитів [16]. Безпосередньо для аналізу текстового вмісту електронних листів застосовуються засоби обробки природної мови, зокрема бібліотека spaCy, яка надає готові інструменти для токенізації, визначення мови, частиномовного аналізу та вилучення іменованих сутностей. Використання таких бібліотек дозволяє зосередитись на логіці предметної області та інтеграції з CRM, не витрачаючи ресурси на розробку низькорівневих алгоритмів обробки тексту з нуля [15], [24]. Сукупне застосування Python, FastAPI та spaCy формує технологічну основу для створення автономного AI-агента, який може бути розгорнутий як окремий сервіс і використовуватись у корпоративному середовищі без порушення цілісності SAP-ландшафту.

1.4 Постановка задачі дослідження та формулювання вимог до системи

На основі проведеного аналізу процесів обробки неструктурованих даних у CRM-системах, а також огляду існуючих методів і інструментальних засобів автоматизації, можна сформулювати загальну науково-практичну проблему дослідження. Вона полягає у відсутності універсального та керованого механізму автоматизованої інтерпретації електронної кореспонденції, здатного забезпечити стабільне перетворення текстових повідомлень у структуровані дані, придатні для використання в бізнес-процесах CRM-систем. Існуючі підходи або не забезпечують достатньої гнучкості (rule-based рішення), або потребують складного налаштування та значних обсягів даних (статистичні ML-моделі), або мають обмеження щодо промислового використання у корпоративному середовищі (LLM без інтеграційної оркестрації).

У межах даної роботи досліджується задача створення інтелектуального програмного агента, який здатний автономно аналізувати вхідні електронні листи, визначати їхню бізнес-релевантність та витягувати ключові атрибути, необхідні для ініціації комерційних процесів у CRM-системі. З формальної точки зору задача зводиться до поєднання методів класифікації текстів, вилучення іменованих сутностей та нормалізації даних у межах єдиного керованого процесу обробки повідомлень.

Відповідно до поставленої задачі до розроблюваної системи висуваються такі функціональні вимоги: автоматичне приймання текстового повідомлення у структурованому форматі; визначення релевантності листа з точки зору комерційного запиту; вилучення

ключових інформаційних атрибутів (назва компанії, контактна особа, товарні позиції, кількісні показники); формування структурованого результату для подальшої інтеграції з CRM-системою.

Окрім функціональних, важливими є нефункціональні вимоги, зокрема вимоги до інтегрованості, масштабованості та керованості рішення. Система повинна бути реалізована у вигляді окремого сервісу, здатного взаємодіяти з CRM через інтеграційну платформу, підтримувати обробку потоків повідомлень у реальному часі та забезпечувати можливість подальшого розширення алгоритмічної частини без зміни інтеграційного контуру.

Таким чином, у межах дослідження ставиться задача розробки та експериментальної перевірки прототипу інтелектуального агента для автоматизації обробки електронної кореспонденції, який поєднує класичні методи машинного навчання та сучасні підходи NLP і може бути використаний у корпоративному CRM-середовищі.

Висновки до розділу 1

У першому розділі виконано аналіз стану проблеми автоматизації обробки неструктурованих текстових даних у корпоративних CRM-системах. Показано, що електронна кореспонденція є одним з основних джерел бізнес-інформації, однак її неструктурований характер суттєво ускладнює використання стандартних засобів автоматизації.

Розглянуто основні підходи до обробки текстової інформації, включаючи rule-based методи, статистичні моделі машинного навчання та сучасні інтелектуальні агентні системи. Встановлено, що

жоден із підходів у чистому вигляді не забезпечує повного задоволення вимог корпоративного середовища, що обумовлює доцільність використання комбінованих рішень.

Проаналізовано інструментальні засоби реалізації системи та обґрунтовано доцільність використання інтеграційної платформи SAP CPI разом із зовнішнім AI-агентом, реалізованим із застосуванням сучасних засобів розробки. На основі проведеного аналізу сформульовано задачу дослідження та визначено основні вимоги до розроблюваної системи, що створює основу для подальшого математичного моделювання та програмної реалізації, які розглядаються у наступних розділах роботи.

2 РОЗРОБКА МОДЕЛЕЙ ТА МЕТОДІВ АВТОМАТИЗАЦІЇ ОБРОБКИ ВХІДНИХ ЕЛЕКТРОННИХ ЛИСТІВ З ВИКОРИСТАННЯМ АГЕНТІВ ШТУЧНОГО ІНТЕЛЕКТУ

2.1 Формалізація задачі автоматизованої обробки вхідного потоку повідомлень

Процес автоматизованої обробки вхідного потоку електронних повідомлень у корпоративних інформаційних системах може бути формалізований як задача інтелектуального аналізу неструктурованих текстових даних з подальшим перетворенням їх у структуроване подання, придатне для використання в бізнес-процесах CRM-систем. Вхідним впливом у такій задачі виступає множина електронних листів, що надходять від зовнішніх контрагентів і характеризуються довільною структурою, багатомовністю та варіативністю лексико-синтаксичних конструкцій. Вихідним результатом є формалізований набір атрибутів, який відображає семантично значущу інформацію повідомлення та може бути безпосередньо використаний для створення або актуалізації бізнес-об’єктів типу Opportunity у CRM-системі. Така постановка задачі узгоджується з класичними підходами до аналізу текстів природною мовою, де ключовим етапом є перехід від неструктурованого тексту до машинозчитуваного подання на основі математичних моделей та методів NLP [3], [6].

Формалізація задачі автоматизованої обробки вхідного потоку повідомлень ґрунтується на поданні процесу у вигляді відображення між множинами вхідних та вихідних даних. Нехай множина вхідних

електронних повідомлень задається як $X = x_1, x_2, \dots, x_n$, де кожен елемент x_i є текстовим повідомленням природною мовою без наперед визначеної структури.

Метою інтелектуальної обробки є побудова відображення F : $X \to Y$, де $Y = y_1, y_2, \dots, y_n$ - множина структурованих результатів аналізу повідомлень. Кожен результат обробки описується у вигляді впорядкованого кортежу $y_i = \langle c_i, a_i \rangle$, у якому змінна $c_i \in 0{,}1$ визначає клас повідомлення за критерієм релевантності (нерелевантне або релевантне), а множина a_i відповідає сукупності семантичних атрибутів, витягнутих із тексту повідомлення, зокрема мови повідомлення, назви компанії, контактної особи, переліку товарних позицій, кількісних характеристик та агрегованих параметрів замовлення (табл. 2.1).

Така формалізація дозволяє чітко розмежувати етапи прийняття рішення щодо релевантності повідомлення та подальшого вилучення інформації, що відповідає класичним підходам до задач автоматичної класифікації текстів і інформаційного вилучення [4], [5], [20].

З урахуванням особливостей вхідних текстових даних та вимог до результатів обробки загальну задачу автоматизованого аналізу повідомлень доцільно декомпозувати на послідовність взаємопов’язаних підзадач. На першому етапі реалізується задача класифікації намірів повідомлення (Intent Classification), метою якої є визначення значення змінної c_i та відсів нерелевантних повідомлень, що не мають бізнес-цінності для процесів формування комерційних можливостей. На другому етапі, який виконується виключно для повідомлень з c_i = 1, здійснюється задача інтелектуального вилучення інформації (Information Extraction), спрямована на формування множини атрибутів a_i. Така двоетапна організація процесу дозволяє зменшити обчислювальні витрати, підвищити стійкість системи до

шумових даних і забезпечити масштабованість рішення, що відповідає рекомендаціям з проєктування інтелектуальних систем обробки текстової інформації в корпоративних середовищах [3], [6], [29]. Формалізована декомпозиція задачі створює методологічну основу для подальшого розгляду математичної моделі класифікації намірів у підрозділі 2.2 та методів вилучення іменованих сутностей і нормалізації атрибутів у підрозділі 2.3.

Таблиця 2.1 – Формалізація елементів математичної моделі задачі автоматизованої обробки повідомлень

Позначення	**Опис елемента моделі**
X	Множина вхідних електронних повідомлень, що надходять до системи інтелектуальної обробки
x_i	Окреме електронне повідомлення природною мовою без наперед визначеної структури
F	Відображення інтелектуальної обробки повідомлень у структуроване подання
Y	Множина структурованих результатів аналізу вхідних повідомлень
y_i	Структурований результат обробки повідомлення x_i
c_i	Ознака релевантності повідомлення (0 - нерелевантне, 1 - релевантне)
a_i	Сукупність семантичних атрибутів, витягнутих із тексту повідомлення
language_i	Визначена мова електронного повідомлення
company_i	Назва компанії-відправника або контрагента
contact_i	Контактна особа, ідентифікована у тексті повідомлення
items_i	Перелік товарних позицій, зазначених у повідомленні
weight_i	Агрегована вага або кількісні характеристики замовлення
i	Індекс повідомлення у множині X, де $i = 1, 2, \ldots, n$

2.2 Математична модель класифікації намірів (Intent Classification) на основі логістичної регресії

Задача класифікації намірів електронних повідомлень у межах досліджуваної системи формалізується як задача бінарної класифікації текстових документів, основною метою якої є автоматизоване визначення належності кожного повідомлення до класу релевантних або нерелевантних з точки зору подальшої обробки в корпоративній CRM-системі. У практичному бізнес-контексті така класифікація забезпечує відокремлення комерційно значущих звернень клієнтів, що потенційно можуть призвести до створення бізнес-об'єктів типу *Opportunity*, від загального потоку електронної кореспонденції, який включає інформаційні розсилки, службові повідомлення, автоматичні сповіщення та спам. Формально кожне вхідне електронне повідомлення позначається як x_i, а результат його аналізу задається бінарною змінною $c_i \in \{0,1\}$, де значення $\mathrm{c_i} = 1$ відповідає релевантному повідомленню, а $c_i = 0$ - нерелевантному. Така постановка задачі дозволяє звести проблему аналізу неструктурованих текстових даних природною мовою до формалізованої математичної моделі прийняття рішень, що створює необхідну основу для застосування методів машинного навчання в інтегрованих корпоративних інформаційних системах та забезпечує масштабованість рішення при обробці великих обсягів вхідної електронної кореспонденції.

Після надходження вхідного електронного повідомлення x_i до системи інтелектуальної обробки воно проходить етапи попередньої обробки тексту, які включають нормалізацію регістру, видалення службових символів, пунктуації та стоп-слів, а також лематизацію або

стемінг залежно від мови повідомлення. Метою цих процедур є зменшення лексичної варіативності тексту та приведення повідомлень до уніфікованого вигляду, що підвищує стійкість подальших математичних моделей до шумових даних. Після завершення попередньої обробки кожне повідомлення x_i перетворюється у числове подання у вигляді вектора ознак v_i, який належить простору R^m, тобто $v_i \in R^m$, де параметр m визначає розмірність простору ознак. Простір ознак формується на основі статистичного подання текстових даних із використанням методу TF-IDF, який дозволяє врахувати як частоту появи термів у конкретному повідомленні, так і їхню інформативність у межах усього корпусу електронних листів. Таке векторне подання забезпечує перехід від неструктурованого тексту природною мовою до формалізованої числової моделі, придатної для застосування методів машинного навчання, а також дозволяє ефективно працювати з великими обсягами електронної кореспонденції за рахунок розрідженості векторів v_i. Використання подання $v_i \in R^m$ створює основу для побудови функції прийняття рішення та подальшої реалізації математичної моделі класифікації намірів повідомлень.

У сформованому просторі ознак задача класифікації намірів електронних повідомлень зводиться до побудови функції прийняття рішення $f(v_i)$, яка відображає вектор ознак v_i у значення бінарної змінної класу c_i. Формально змінна класу належить множині $\{0,1\}$, тобто $c_i \in \{0,1\}$, де значення $c_i = 1$ інтерпретується як наявність у повідомленні комерційно значущого наміру, а $c_i = 0$ - як відсутність такого наміру. Така постановка задачі дозволяє чітко формалізувати процес прийняття рішення та відокремити релевантні повідомлення від нерелевантних на основі їхнього векторного подання.

Функція прийняття рішення $f(v_i)$ будується таким чином, щоб максимально точно апроксимувати залежність між вхідним вектором ознак та істинною міткою класу повідомлення, яка задається змінною $y_i \in \{0,1\}$. У процесі навчання моделі значення y_i використовуються як еталонні відповіді для коригування параметрів класифікатора, що дозволяє мінімізувати кількість помилкових рішень. Важливою особливістю задачі є те, що функція $f(v_i)$ має працювати у просторі великої розмірності m, що зумовлено використанням статистичних методів подання тексту. Тому при виборі математичної моделі необхідно враховувати її здатність ефективно працювати з розрідженими векторами ознак та забезпечувати стабільну якість класифікації навіть за наявності шуму та неоднорідності текстових даних.

Саме з огляду на ці вимоги у роботі обрано модель логістичної регресії як функцію прийняття рішення $f(v_i)$, оскільки вона поєднує відносну простоту реалізації, інтерпретованість результатів і доведену ефективність у задачах бінарної класифікації текстів. Логістична регресія дозволяє не лише визначати клас повідомлення, але й отримувати кількісну оцінку впевненості моделі у прийнятому рішенні, що є важливим для подальшого використання результатів класифікації в корпоративних інформаційних системах.

Математична модель логістичної регресії, що використовується для класифікації намірів електронних повідомлень, ґрунтується на побудові лінійної комбінації компонентів вектора ознак v_i з подальшим нелінійним перетворенням отриманого значення. Для кожного повідомлення обчислюється скалярна величина z_i, яка визначається виразом $z_i = w^T v_i + b$, де $w \in R^m$ є вектором вагових коефіцієнтів моделі, а b - параметром зсуву, що дозволяє зміщувати поріг прийняття рішення. Значення z_i відображає узагальнену міру

відповідності повідомлення класу релевантних з урахуванням внеску кожної ознаки у результат класифікації.

Для перетворення значення z_i у ймовірнісну оцінку використовується сигмоїдна функція, яка забезпечує відображення дійсної осі у відрізок $[0,1]$. Ймовірність належності повідомлення x_i до класу релевантних визначається як $P(c_i = 1 \mid v_i) = \frac{1}{1+e^{-z_i}}$. Така ймовірнісна інтерпретація результату дозволяє не лише отримати бінарне рішення щодо класу повідомлення, але й оцінити ступінь упевненості моделі у цьому рішенні, що є важливим для корпоративних систем підтримки прийняття рішень. У практичних сценаріях це дає змогу гнучко налаштовувати поведінку системи залежно від бізнес-вимог, зокрема змінювати порогове значення прийняття рішення.

Остаточне віднесення повідомлення до одного з класів здійснюється на основі порогового правила, згідно з яким повідомлення вважається релевантним у випадку виконання умови $P(c_i = 1 \mid v_i) \geq \tau$, де τ є заданим пороговим значенням. Якщо зазначена умова не виконується, тобто $P(c_i = 1 \mid v_i) < \tau$, повідомлення класифікується як нерелевантне. У більшості практичних реалізацій параметр порогу обирається рівним $\tau = 0.5$, що забезпечує симетричний баланс між помилками першого та другого роду. Таким чином, логістична регресія реалізує гнучкий і інтерпретований механізм прийняття рішень, який добре масштабується для обробки великих масивів текстових даних і є придатним для використання в автоматизованих системах аналізу електронної кореспонденції.

Навчання моделі логістичної регресії у межах задачі класифікації намірів електронних повідомлень зводиться до задачі мінімізації функції логістичних втрат, яка кількісно характеризує

відхилення прогнозованих моделлю значень від істинних міток класу навчальної вибірки. Формально для кожного повідомлення x_i задається істинна мітка класу $y_i \in \{0,1\}$, а прогнозоване значення моделі визначається як $h_\beta(x_i)$, що інтерпретується як ймовірність належності повідомлення до класу релевантних. Для оцінювання сумарної помилки моделі на навчальній вибірці використовується функція логістичних втрат (1), або *Binary Cross-Entropy*, яка враховує ймовірнісний характер вихідних значень класифікатора.

Функція логістичних втрат визначається наступним чином:

$$J(\beta) = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \ln\left(h_\beta(x_i)\right) + (1-y_i)\ln\left(1-h_\beta(x_i)\right)\right], \quad (2.1)$$

де:

$J(\beta)$ - значення цільової функції оптимізації

β - вектор параметрів моделі логістичної регресії

N - кількість повідомлень у навчальній вибірці

y_i - істинна мітка класу

$h_\beta(x_i)$ - прогнозована моделлю ймовірність релевантності повідомлення.

Така форма функції втрат забезпечує суттєве штрафування впевнених, але помилкових рішень моделі, що є критично важливим у задачах аналізу текстів з високим рівнем варіативності та шуму.

Мінімізація функції $J(\beta)$ здійснюється шляхом ітеративного коригування параметрів моделі β із використанням градієнтних методів оптимізації. На кожній ітерації навчання обчислюється градієнт функції втрат за параметрами моделі, що дозволяє визначити напрямок і величину їх зміни з метою зменшення значення цільової функції. Такий підхід забезпечує збіжність алгоритму до

локального мінімуму та дозволяє ефективно навчати модель навіть за умов високої розмірності простору ознак і розрідженості векторів v_i. Застосування логістичної регресії у поєднанні з TF-IDF-поданням текстів забезпечує стабільну якість класифікації та робить модель придатною для використання в промислових системах автоматизованої обробки електронної кореспонденції.

Вибір моделі логістичної регресії для розв'язання задачі класифікації намірів електронних повідомлень є обґрунтованим як з математичної, так і з практичної точки зору. Однією з ключових переваг даної моделі є її інтерпретованість, яка дозволяє аналізувати вплив окремих компонентів вектора ознак v_i на результат класифікації через значення вагових коефіцієнтів w. На відміну від більш складних моделей машинного навчання, логістична регресія забезпечує прозорий механізм прийняття рішень, що є важливим у корпоративних інформаційних системах, де результати автоматизованого аналізу повинні бути зрозумілими для аналітиків і бізнес-користувачів.

Ймовірнісна природа вихідного значення моделі у вигляді $h_\beta(x_i)$ дозволяє використовувати результат класифікації не лише як жорстке бінарне рішення, але і як кількісну міру впевненості у наявності комерційного наміру в електронному повідомленні. Це, у свою чергу, створює можливості для гнучкого налаштування порогового значення τ залежно від бізнес-вимог, наприклад, зменшення кількості хибнопозитивних спрацьовувань або мінімізації ризику пропуску реальних комерційних запитів. Такий підхід дозволяє адаптувати поведінку системи до специфіки предметної області та характеру вхідного потоку повідомлень.

Крім того, логістична регресія добре масштабується для роботи з високорозмірними та розрідженими просторами ознак, що є характерним для текстових даних, поданих у вигляді TF-IDF-векторів

$v_i \in R^m$. Поєднання відносно невеликої обчислювальної складності, стійкості до шуму та стабільної якості класифікації робить дану модель придатною для використання в інтегрованих системах автоматизованої обробки електронної кореспонденції. Таким чином, застосування логістичної регресії у межах досліджуваної системи забезпечує ефективне розв'язання задачі класифікації намірів і створює надійну основу для подальших етапів інтелектуальної обробки даних.

2.3 Метод вилучення іменованих сутностей (NER) та модель нормалізації атрибутів

Після визначення релевантності електронного повідомлення наступним етапом інтелектуальної обробки є вилучення з тексту семантично значущих атрибутів, необхідних для формування структурованого опису бізнес-запиту. Дана задача формалізується як задача розпізнавання іменованих сутностей (*Named Entity Recognition, NER*), метою якої є ідентифікація у неструктурованому тексті фрагментів, що відповідають наперед визначеним класам сутностей. У межах досліджуваної системи NER використовується для визначення таких атрибутів, як назва компанії-контрагента, контактна особа, а також товарні позиції і кількісні характеристики замовлення. Застосування методів NER дозволяє перейти від суцільного тексту електронного листа до набору формалізованих елементів, придатних для подальшої нормалізації та інтеграції в корпоративні інформаційні системи [3], [6] (табл. 2.2).

Таблиця 2.2 – Типи іменованих сутностей, що вилучаються з електронного повідомлення

Тип сутності	**Опис**
ORG	Назва компанії або організації-контрагента
PERSON	Контактна особа, зазначена у повідомленні
PRODUCT	Найменування товарної позиції
QUANTITY	Кількісна характеристика товару
UNIT	Одиниця виміру (т, кг, шт тощо)
WEIGHT	Узагальнена вага або маса замовлення
LANGUAGE	Мова тексту електронного повідомлення

Для підвищення точності вилучення атрибутів у роботі застосовано гібридний підхід, що поєднує статистичні методи NER та правило-орієнтовану нормалізацію даних. Моделі розпізнавання іменованих сутностей використовуються для ідентифікації загальних лінгвістичних об’єктів, таких як організації та імена осіб, тоді як вилучення товарних позицій, кількісних параметрів і одиниць виміру реалізується з використанням регулярних виразів і доменно-орієнтованих правил. Такий підхід дозволяє враховувати специфіку предметної області та зменшити кількість помилок, пов’язаних із варіативністю форм запису технічних і комерційних даних у діловому листуванні [4], [7].

2.4 Алгоритм функціонування мультиагентної системи обробки даних

Функціонування розробленої системи автоматизованої обробки вхідних електронних повідомлень реалізується у вигляді мультиагентного алгоритму, в якому окремі програмні компоненти

виконують спеціалізовані функції аналізу, інтерпретації та обробки даних. Такий підхід дозволяє розглядати систему як сукупність взаємодіючих агентів, кожен з яких відповідає за окремий етап обробки інформації та реалізує чітко визначену функціональну роль. Загальний алгоритм роботи системи ґрунтується на послідовному проходженні електронного повідомлення через етапи сприйняття, семантичної інтерпретації, прийняття рішення щодо релевантності та формування структурованого результату, що відповідає класичній моделі інтелектуального агента.

Координацію взаємодії між агентами, а також обмін даними з зовнішніми корпоративними інформаційними системами виконує інтеграційний контур. Він забезпечує надійну передачу повідомлень між компонентами системи, контроль поточного стану обробки та централізовану обробку результатів аналізу. Така організація архітектури дозволяє розмежувати відповідальність між окремими програмними модулями, підвищити масштабованість рішення, а також спростити його подальше розширення та адаптацію до нових бізнес-вимог без суттєвих змін загальної логіки роботи системи [3], [11].

Алгоритм функціонування мультиагентної системи обробки електронних повідомлень складається з наступних послідовних етапів:

1. Отримання електронного повідомлення інтеграційною платформою та формування його структурованого подання у форматі JSON для подальшої автоматизованої обробки.

2. Передача вхідного повідомлення до інтелектуального агента, який ініціює процес аналізу текстових даних.

3. Виконання етапу класифікації наміру повідомлення з використанням моделі логістичної регресії з метою визначення його релевантності.

4. Перевірка результату класифікації та прийняття рішення щодо подальшої обробки повідомлення залежно від визначеного класу.

5. У разі визначення повідомлення як релевантного - запуск етапу вилучення іменованих сутностей та нормалізації атрибутів, необхідних для бізнес-аналізу.

6. Формування структурованого результату аналізу у вигляді набору атрибутів, які використовуються для створення або оновлення бізнес-об’єкта.

7. Повернення результату обробки до інтеграційного контуру для подальшої взаємодії з корпоративними системами.

8. Ініціація створення або оновлення об’єкта Opportunity у CRM-системі на основі отриманих структурованих даних.

9. Завершення циклу обробки повідомлення та фіксація результатів виконання для подальшого аналізу та контролю.

Запропонований алгоритм реалізує ієрархічну модель обробки повідомлень, у якій рішення про подальші дії приймається на основі результатів попередніх етапів аналізу. Такий підхід дозволяє уникнути надлишкових обчислень для нерелевантних повідомлень, зменшити навантаження на обчислювальні ресурси та забезпечити ефективну обробку великих обсягів вхідної електронної кореспонденції. Крім того, мультиагентна організація алгоритму спрощує модифікацію та розширення системи за рахунок додавання нових агентів або заміни окремих моделей без необхідності зміни загальної логіки функціонування рішення [11], [29].

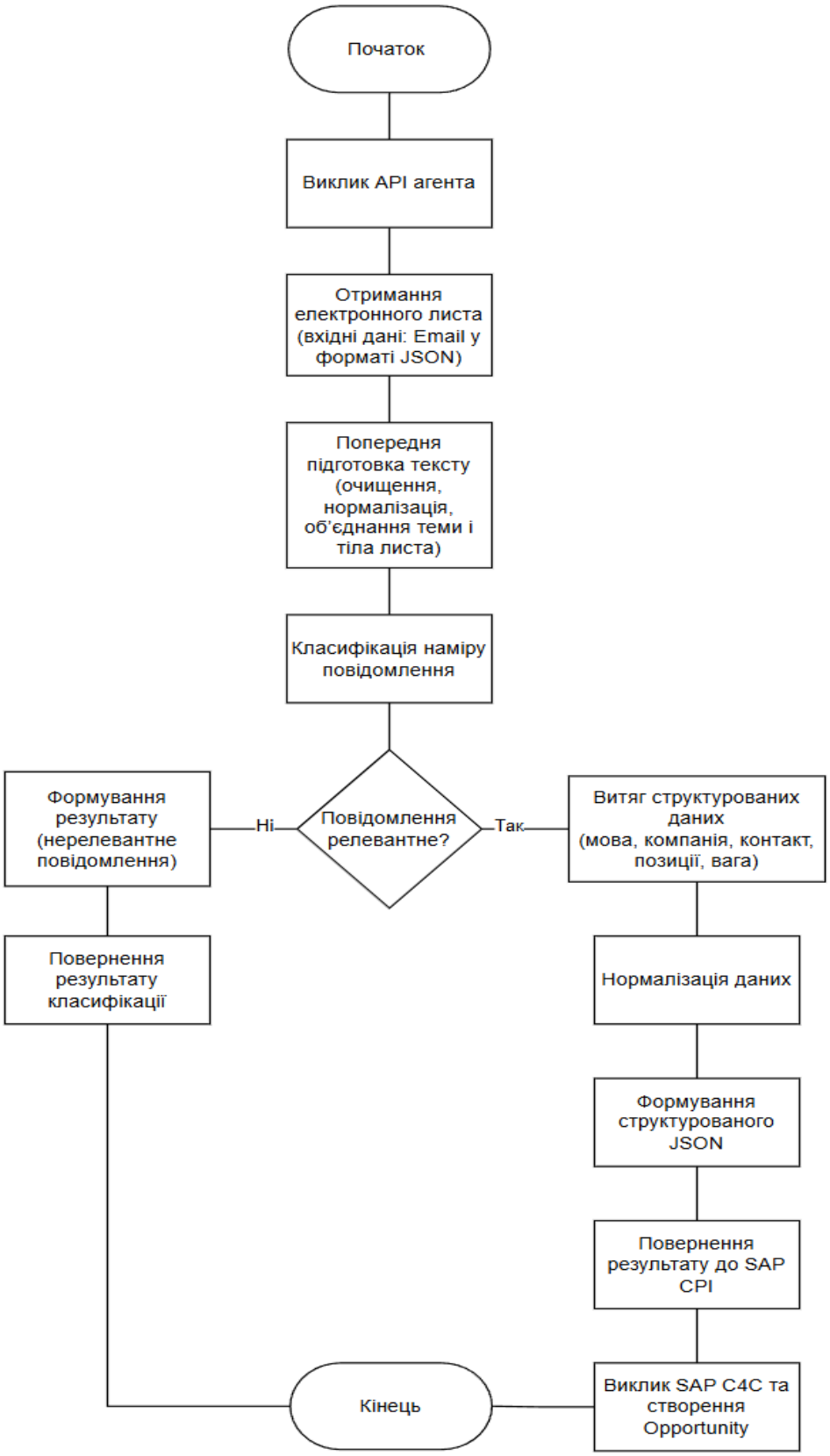

Рисунок 2.1 - Функціональна схема розроблюваної програмної системи

2.5 Вибір критеріїв якості та метрик оцінювання ефективності моделей

Оцінювання ефективності розроблених моделей інтелектуальної обробки електронних повідомлень здійснюється на основі стандартних метрик якості бінарної класифікації та інформаційного вилучення, що широко застосовуються у задачах аналізу текстів і машинного навчання. Вибір критеріїв оцінювання обумовлений необхідністю отримання об’єктивної кількісної оцінки здатності моделей коректно визначати релевантні повідомлення у потоці вхідної електронної кореспонденції, а також мінімізувати кількість помилок першого та другого роду. Зазначені помилки мають різний вплив на бізнес-процеси, тому їх урахування є принципово важливим для практичного використання результатів моделювання.

Для задачі класифікації намірів ключовими є метрики точності та повноти, які дозволяють оцінити баланс між кількістю правильно ідентифікованих релевантних повідомлень і загальною кількістю спрацьовувань моделі. У контексті корпоративних CRM-систем особливе значення має зменшення кількості хибнопозитивних результатів, оскільки помилкове створення бізнес-об’єктів типу *Opportunity* призводить до зайвого навантаження на користувачів системи, зниження довіри до автоматизованих рішень та спотворення аналітичних і звітних даних [3], [6]. Для оцінювання цього аспекту використовується метрика точності (*Precision*) (2), яка визначається як відношення кількості правильно класифікованих релевантних повідомлень до загальної кількості повідомлень, які модель віднесла до релевантних за формулою 2.2.

$$Precision = \frac{TP}{(TP + FP)} \quad (2.2)$$

де:

- TP (True Positives) - кількість правильно класифікованих замовлень;
- FP (False Positives) - кількість помилково класифікованого спаму.

Метрика повноти (*Recall*) (3) відображає здатність інтелектуального агента знаходити всі реальні комерційні запити у потоці вхідної електронної кореспонденції та характеризує частку виявлених релевантних повідомлень серед усіх фактично релевантних звернень. Вона визначається наступним співвідношенням на формулі 2.3.

$$Recall = \frac{TP}{(TP + FN)} \quad (2.3)$$

де FN (*False Negatives*) - кількість релевантних повідомлень, помилково класифікованих як нерелевантні.

У бізнес-контексті такі помилки відповідають пропущеним замовленням або втраченим можливостям продажу, що є критичними для процесу комерційної діяльності та безпосередньо впливають на фінансові показники компанії.

З огляду на те, що для бізнес-процесу продажів однаково критичними є як пропуск реального замовлення, так і створення хибної комерційної можливості (*Opportunity*), у роботі використовується інтегральна метрика F1-score (4), яка є гармонійним середнім між показниками точності та повноти та дозволяє отримати узагальнену оцінку якості моделі за формулою 2.4.

$$F_1 = 2 * \frac{(Precision * Recall)}{(Precision + Recall)} \quad (2.4)$$

Застосування метрики *F1-score* дозволяє забезпечити збалансовану оцінку ефективності моделі класифікації намірів електронних повідомлень, враховуючи як хибнопозитивні, так і хибнонегативні помилки. Такий підхід є загальноприйнятим у задачах аналізу текстів і машинного навчання та забезпечує коректну основу для порівняння різних моделей у межах експериментального дослідження [5], [20].

Висновки до розділу 2

У другому розділі магістерської роботи виконано математичне моделювання процесів інтелектуальної обробки вхідних електронних повідомлень у корпоративних інформаційних системах. Розглянуто формалізацію задачі автоматизованої обробки текстових даних, що дозволило подати вхідний потік електронної кореспонденції у вигляді множини повідомлень та визначити вимоги до структурованого результату аналізу. Така формалізація створила основу для подальшого вибору методів машинного навчання та алгоритмів обробки природної мови.

У межах розділу обґрунтовано вибір логістичної регресії як базової математичної моделі для класифікації намірів електронних повідомлень. Показано, що дана модель є доцільною для задач бінарної класифікації текстів великої розмірності, сформованих із використанням TF-IDF-представлення, та забезпечує

інтерпретованість результатів у вигляді ймовірнісних оцінок. Описано процес навчання моделі шляхом мінімізації функції логістичних втрат, що дозволяє формалізувати задачу оптимізації параметрів та забезпечити стійку роботу класифікатора в умовах шумових і варіативних текстових даних.

Окрему увагу приділено методам вилучення іменованих сутностей та нормалізації атрибутів електронних повідомлень. Запропоновано гібридний підхід, що поєднує статистичні методи NER і правило-орієнтовані алгоритми, орієнтовані на специфіку предметної області. Такий підхід дозволяє підвищити точність ідентифікації бізнес-значущих атрибутів, зокрема товарних позицій, кількісних характеристик та одиниць виміру, що є критично важливим для подальшої інтеграції результатів аналізу в CRM-систему.

У розділі також розроблено алгоритм функціонування мультиагентної системи обробки даних, який реалізує послідовний цикл аналізу повідомлення - від отримання вхідних даних до створення комерційної можливості у CRM-системі. Запропонований алгоритм забезпечує декомпозицію задачі на окремі етапи, дозволяє оптимізувати використання обчислювальних ресурсів та спрощує масштабування рішення за рахунок модульної архітектури.

Для оцінювання ефективності розроблених моделей визначено систему критеріїв якості та метрик оцінювання, що базуються на показниках точності, повноти та інтегральної метрики F1-score. Обґрунтовано доцільність використання зазначених метрик у контексті корпоративних CRM-систем, де критичними є як пропуск реальних комерційних запитів, так і створення хибних бізнес-об’єктів. Застосування цих критеріїв формує об’єктивну основу для експериментального дослідження та порівняння ефективності моделей, що буде виконано у наступному розділі роботи.

3 РОЗРОБКА ПРОГРАМНОГО КОМПЛЕКСУ AI EMAIL PARSER AGENT

3.1 Розробка бізнес-процесів автоматизації створення об’єктів CRM на основі інтелектуальної обробки вхідних електронних повідомлень

Під час цифровізації бізнес-процесів у великих організаціях, зберігається низка проблем, пов’язаних із ручною та напівавтоматичною обробкою вхідної електронної кореспонденції. Значна частина комерційно важливої інформації, включаючи запити на розрахунок вартості продукції, уточнення умов постачання та ініціацію потенційних угод, надходить до підприємства у вигляді неструктурованих текстових повідомлень природною мовою. Такі повідомлення потребують попередньої інтерпретації, аналізу змісту та подальшого перенесення релевантних даних до CRM-систем для формування або актуалізації бізнес-об’єктів. За відсутності спеціалізованого програмного забезпечення зазначені операції виконуються переважно вручну, що призводить до зростання навантаження на персонал, збільшення часу обробки клієнтських запитів та підвищення ймовірності помилок, обумовлених людським фактором [29, 31].

Однією з ключових проблем є несистемна обробка комерційних запитів, що надходять електронною поштою. Частина релевантних листів може бути оброблена із затримкою або втрачена серед загального потоку кореспонденції, тоді як нерелевантні повідомлення витрачають час співробітників відділу продажів. Іншою суттєвою

проблемою є відсутність уніфікованого підходу до структурування даних, отриманих з електронних листів, що призводить до неповноти або некоректності інформації в CRM-системі. Крім того, ручна інтерпретація тексту значною мірою залежить від людського фактору та суб'єктивного досвіду конкретного працівника, що негативно впливає на відтворюваність бізнес-процесів і якість управлінських рішень [18, 29].

Вирішення зазначених проблем пропонується здійснювати шляхом розробки та впровадження програмного комплексу AI Email Parser Agent, який реалізує автоматизовану класифікацію електронних листів, витяг структурованих даних та інтеграцію результатів аналізу з CRM-системою відповідно до сучасних підходів упровадження штучного інтелекту в корпоративні інформаційні системи [29, 31]. Основні проблеми та можливі шляхи їх вирішення наведені у таблиці 3.1.

Таблиця 3.1 – Можливі варіанти вирішення проблем

Проблема	Ручна обробка вхідних електронних листів, що містять комерційні запити
Стосується	Менеджера з продажів, аналітика CRM-системи
Її наслідком є	Значні витрати часу на аналіз кореспонденції, затримки у створенні комерційних можливостей, підвищений ризик втрати релевантних запитів
Успішне рішення	Автоматизація класифікації електронних листів за релевантністю з використанням методів машинного навчання
Проблема	Неструктурованість даних у тексті електронних листів
Стосується	CRM-системи та бізнес-аналітики
Її наслідком є	Неповні або некоректні записи в CRM, складність подальшого аналізу даних і формування звітності
Успішне рішення	Використання інтелектуального агента для автоматичного витягу та нормалізації структурованих атрибутів з текстових повідомлень

Продовження табл. 3.1

Проблема	Залежність якості обробки електронних листів від людського фактору
Стосується	Бізнес-процесів відділу продажів і підтримки клієнтів
Її наслідком є	Суб'єктивність прийняття рішень, помилки при інтерпретації запитів, нерівномірна якість обробки повідомлень
Успішне рішення	Впровадження інтелектуального агентного програмного забезпечення з формалізованою логікою обробки та повторюваними алгоритмами аналізу
Проблема	Затримка передачі інформації з електронних листів до CRM-системи
Стосується	Інтеграційних процесів підприємства
Її наслідком є	Втрата оперативності реагування на запити клієнтів та зниження ефективності бізнес-процесів
Успішне рішення	Інтеграція інтелектуального агента з CRM-системою через інтеграційну платформу SAP Cloud Platform Integration

В свою чергу, в таблиці 3.2 наведено позиціювання програмного продукту, що розробляється, з точки зору його призначення та практичної цінності для підприємства.

Таблиця 3.2 - Визначення позиції програмного продукту

Для	Корпоративного підприємства з великою кількістю вхідної електронної кореспонденції
Якому	Необхідно автоматизувати обробку комерційних запитів та підвищити ефективність роботи CRM
Назва продукту	Прототип інтелектуального агента AI Email Parser Agent для SAP C4C
Який	Орієнтований на ML та NLP, сервісно-орієнтований, інтегрований з CRM
На відміну від	Ручної обробки або простих rule-based фільтрів
Наш продукт	Забезпечує автоматичну класифікацію, витяг даних і створення Opportunity

Основною задачею програмного комплексу є автоматизований аналіз вхідних електронних листів із подальшим формуванням структурованого результату, придатного для використання в бізнес-процесах CRM-системи. Для забезпечення коректної та відтворюваної роботи програмного забезпечення процес його функціонування поділено на низку логічних етапів (рис. 3.1).

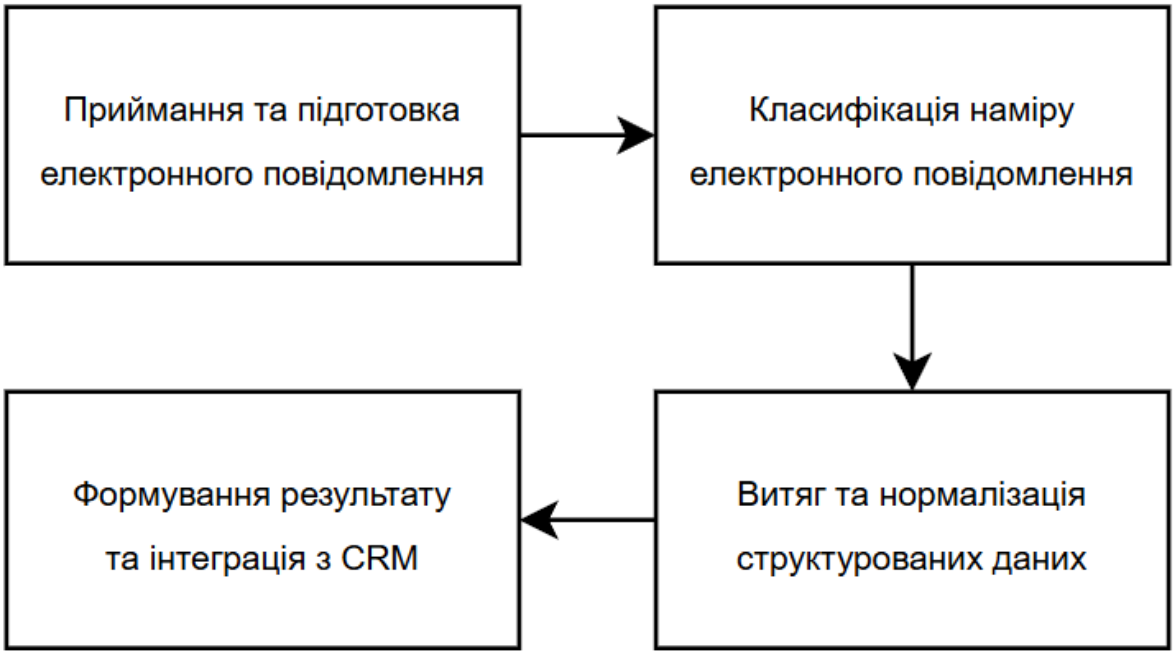

Рисунок 3.1 - Діаграма етапів функціонування програмного комплексу AI Email Parser Agent

До основних етапів функціонування системи належать: прийом електронного листа у форматі JSON; попередня обробка тексту; класифікація листа за релевантністю; витяг структурованих даних; формування вихідного JSON та передача результатів до CRM-системи. Така декомпозиція дозволяє розділити відповідальність між окремими модулями програмного комплексу та спростити подальше розширення функціональності системи.

Важливим аспектом опрацювання програмної частини є також визначення ролей користувачів та системних компонентів, що взаємодіють із програмним агентом. У межах розроблюваного рішення виділяються ролі адміністратора системи, користувачів CRM та інтеграційної платформи, кожна з яких має чітко визначені функції та рівень доступу. Такий підхід відповідає принципам проєктування корпоративних інформаційних систем та забезпечує керованість і безпечність експлуатації програмного продукту.

Розглянемо детально етапи функціонування програмного комплексу AI Email Parser Agent, що забезпечує автоматизовану обробку вхідних електронних повідомлень та інтеграцію результатів аналізу з CRM-системою. Логіка роботи системи побудована у вигляді послідовності взаємопов’язаних етапів, кожен з яких виконує окрему функціональну задачу та формує вхідні дані для наступного етапу.

Першим етапом роботи програмного комплексу є приймання вхідного електронного повідомлення від інтеграційної платформи (рис. 3.2). Повідомлення надходить у структурованому форматі JSON та містить службову інформацію (відправник, одержувач, тема листа), а також основний текст повідомлення.

Рисунок 3.2 – Основні функції етапу «Приймання та підготовка електронного повідомлення»

На даному етапі виконується попередня підготовка текстових даних, яка включає очищення тексту від службових символів, нормалізацію регістру, а також об’єднання тексту теми та тіла листа для подальшого аналізу. В таблиці 3.3 наведені характеристики етапу «Приймання та підготовка електронного повідомлення»

Наступним етапом є класифікація електронного листа за релевантністю, тобто визначення, чи є повідомлення комерційним запитом, що потребує створення запису Opportunity у CRM-системі (рис. 3.3). Даний етап реалізується з використанням методів машинного навчання на основі векторного подання тексту та бінарного класифікатора.

Таблиця 3.3 – Характеристика етапу «Приймання та підготовка електронного повідомлення»

Назва характеристики	Значення характеристики
Ім’я етапу	Приймання та підготовка електронного повідомлення
Основні учасники	Інтеграційна платформа, інтелектуальний агент
Вхідна подія	Надходження електронного листа
Вхідні документи	JSON-представлення електронного повідомлення
Вихідна подія	Передача підготовленого тексту на етап класифікації
Вихідні документи	Нормалізований текст повідомлення
Клієнт етапу	Модуль класифікації

У разі визначення повідомлення як нерелевантного подальша обробка припиняється, а система повертає результат класифікації. Якщо ж повідомлення визначене як релевантне, ініціюється наступний етап аналізу тексту.

В таблиці 3.4 наведені характеристики етапу «Класифікація наміру повідомлення».

Таблиця 3.4 – Характеристика етапу «Класифікація наміру повідомлення»

Назва характеристики	Значення характеристики
Ім’я етапу	Класифікація наміру повідомлення
Основні учасники	Інтелектуальний агент
Вхідна подія	Підготовлений текст повідомлення
Вхідні документи	Нормалізований текст
Вихідна подія	Визначення релевантності
Вихідні документи	Результат класифікації, коефіцієнт впевненості
Клієнт етапу	Модуль витягу даних / інтеграційна платформа

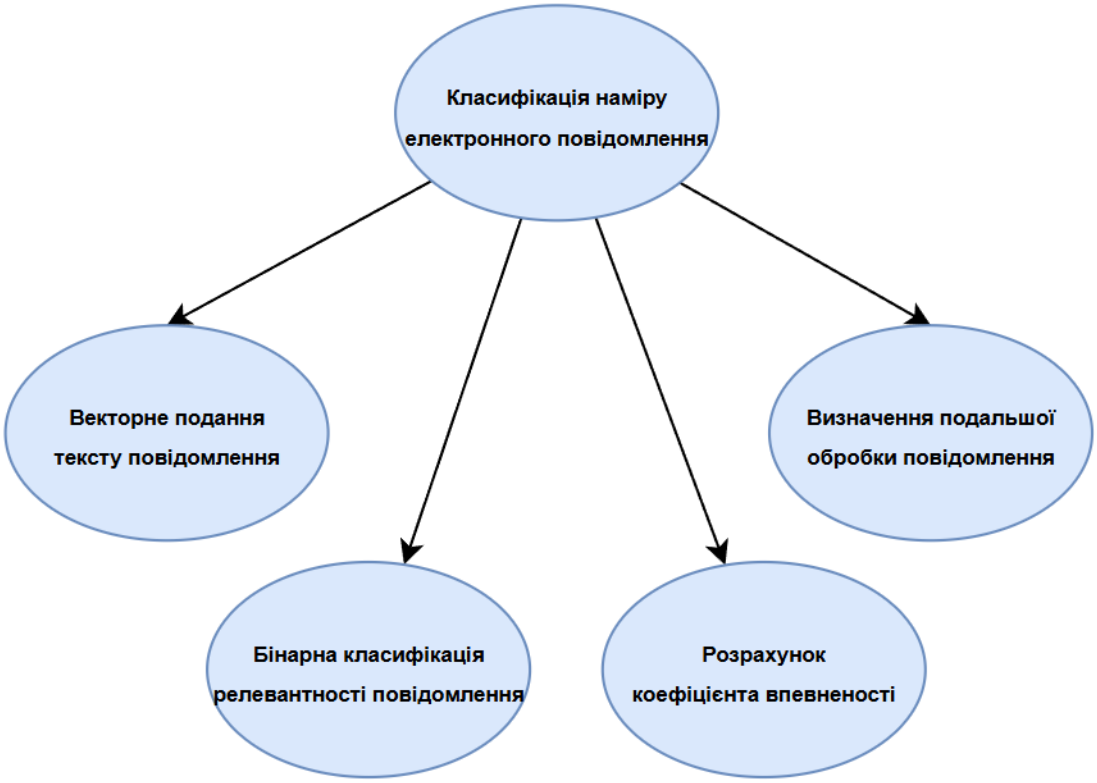

Рисунок 3.3 – Основні функції етапу «Класифікація наміру повідомлення»

У разі позитивного результату класифікації виконується витяг структурованих атрибутів з тексту електронного листа. До таких атрибутів належать мова повідомлення, назва компанії, контактна особа, перелік товарних позицій та загальна вага замовлення (рис. 3.4).

Для реалізації даного етапу використовується поєднання методів автоматичного розпізнавання іменованих сутностей та евристичних правил, що дозволяє досягти високої точності витягу даних у межах предметної області.

В таблиці 3.5 наведені характеристики етапу «Витяг та нормалізація даних».

Рисунок 3.4 - Основні функції етапу «Витяг та нормалізація даних»

Таблиця 3.5 - Характеристика етапу «Витяг та нормалізація даних»

Назва характеристики	Значення характеристики
Ім’я етапу	Витяг та нормалізація даних
Основні учасники	Інтелектуальний агент
Вхідна подія	Релевантне повідомлення
Вхідні документи	Текст повідомлення
Вихідна подія	Формування структурованого результату
Вихідні документи	Структурований JSON
Клієнт етапу	CRM-система

Завершальним етапом є формування структурованого результату аналізу та передача його до CRM-системи через інтеграційну платформу (рис. 3.5). На основі отриманих даних

автоматично створюється запис Opportunity із заповненими ключовими атрибутами.

Рисунок 3.5 - Етап «Формування результату та інтеграція з CRM»

В таблиці 3.6 наведені характеристики етапу «Формування результату та інтеграція».

Таблиця 3.6 - Характеристика етапу «Формування результату та інтеграція»

Назва характеристики	Значення характеристики
Ім’я етапу	Формування результату та інтеграція
Основні учасники	Інтелектуальний агент, CRM
Вхідна подія	Завершення витягу даних
Вхідні документи	Структурований JSON
Вихідна подія	Створення Opportunity
Вихідні документи	Запис у CRM
Клієнт етапу	Бізнес-процеси CRM

Розглянемо відомості про користувачів розроблюваної програмної системи AI Email Parser Agent. Перелік користувачів

системи сформовано з урахуванням ролей, які беруть участь у процесі обробки вхідних електронних повідомлень, аналізу результатів та використання сформованих структурованих даних у бізнес-процесах CRM-системи. До основних користувачів належать: адміністратор системи, керівник підрозділу (начальник відділу продажів), бізнес-користувач (менеджер з продажів) та інтеграційний користувач системи.

Адміністратор системи здійснює контроль усіх дій користувачів, відповідає за стабільність роботи програмного комплексу, усуває помилки, виконує налаштування параметрів системи та забезпечує її коректну взаємодію з інтеграційною платформою та CRM-системою. Керівник підрозділу має наглядові функції та використовує систему переважно для перегляду результатів автоматизованої обробки електронних листів, аналізу сформованих комерційних можливостей та оцінювання ефективності роботи підрозділу. Бізнес-користувач (менеджер з продажів) працює з результатами аналізу електронних повідомлень, перевіряє коректність витягнутих даних, за необхідності ініціює подальші дії у CRM-системі та формує звітні матеріали.

Інтеграційний користувач представляє зовнішню систему або сервіс (інтеграційну платформу), який забезпечує передачу електронних повідомлень до інтелектуального агента та отримання результатів обробки для подальшого використання в CRM-системі. У таблиці 3.7 наведено профілі користувачів розроблюваної програмної системи.

Перелік акторів розроблюваної програмної системи AI Email Parser Agent сформовано з урахуванням ролей, що беруть участь у процесі автоматизованої обробки електронних повідомлень, контролю результатів аналізу та використання сформованих структурованих даних у CRM-системі. До складу акторів системи

належать: адміністратор системи, керівник підрозділу, менеджер з продажів та інтеграційна платформа.

Таблиця 3.7 - Профілі користувачів

Типовий представник	Адміністратор
Опис	Користувач системи, наділений правами адміністратора, має доступ до всіх функцій системи
Тип	Адміністратор
Відповідальність	Забезпечує стабільну роботу системи, усуває помилки, налаштовує інтеграцію з CRM
Критерій успіху	Безперервна та коректна обробка електронних повідомлень
Типовий представник	Керівник підрозділу
Опис	Користувач системи, наділений наглядовими функціями
Тип	Користувач
Відповідальність	Контроль результатів автоматизованої обробки листів
Критерій успіху	Своєчасне отримання аналітичних даних та звітів
Типовий представник	Менеджер з продажів
Опис	Користувач системи, який працює з результатами аналізу електронних листів
Тип	Користувач
Відповідальність	Перевірка та використання структурованих даних у CRM
Критерій успіху	Швидке створення та обробка Opportunity
Типовий представник	Інтеграційний користувач
Опис	Системний користувач, що забезпечує обмін даними між компонентами
Тип	Системний користувач
Відповідальність	Передача та отримання даних між системами
Критерій успіху	Відсутність помилок інтеграції

Адміністратор системи здійснює контроль усіх дій користувачів, відповідає за працездатність та стабільність роботи програмного комплексу, виконує налаштування параметрів аналізу, інтеграційних з'єднань та контролює коректність взаємодії з CRM-системою.

Керівник підрозділу має наглядові функції та використовує систему для ознайомлення з результатами автоматизованої обробки електронних листів, аналізу створених комерційних можливостей та оцінювання ефективності роботи підрозділу.

Менеджер з продажів є основним бізнес-користувачем системи. Він працює з результатами аналізу електронних повідомлень, перевіряє коректність витягнутих даних, використовує сформовані записи Opportunity у CRM-системі та, за необхідності, формує звітні матеріали.

Інтеграційна платформа виступає системним актором, що забезпечує обмін даними між поштовими сервісами, інтелектуальним агентом аналізу тексту та CRM-системою. Даний актор не взаємодіє з користувачем безпосередньо, але є критично важливим для коректної роботи всієї системи (рис. 3.6).

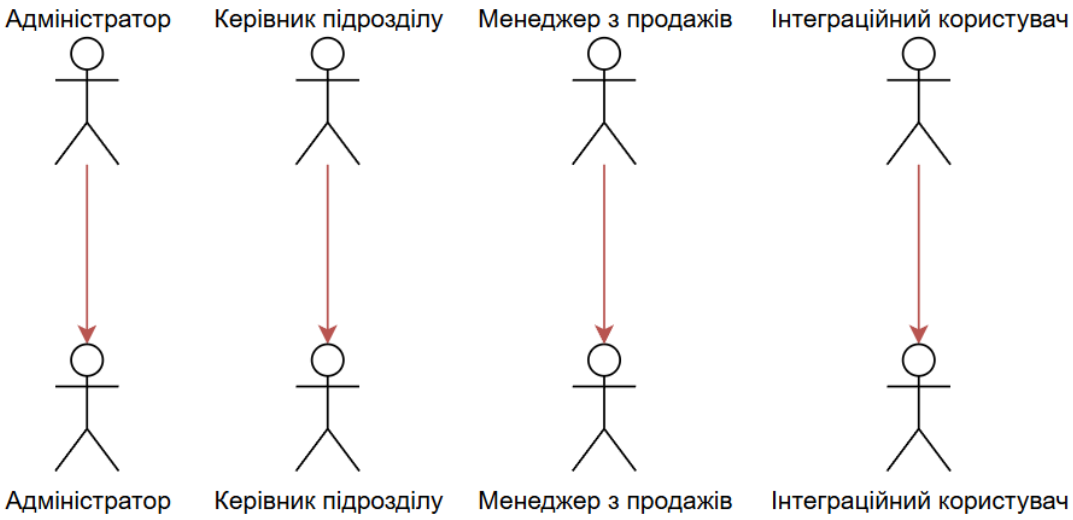

Рисунок 3.6 - Аналіз акторів системи

В таблиці 3.8 наведено короткий опис наявних у системі акторів для ПЗ «AI Email Parser Agent». Для кожного з визначених акторів було виявлено основні варіанти використання програмної системи. Кожному варіанту використання надано коротке формулювання, що відображає його призначення. Виявлені варіанти використання зведені у таблицю 3.9.

Таблиця 3.8 - Короткий опис акторів

Актор	Короткий опис
Адміністратор	Користувач системи, наділений правами адміністратора; має доступ до всіх функцій, виконує налаштування сервісу, моніторинг, керування конфігураціями та журналами обробки
Менеджер з продажів	Користувач системи, який переглядає результати аналізу електронних листів, перевіряє витягнуті атрибути та використовує сформовані дані для подальшої роботи з Opportunity у CRM
Керівник підрозділу	Користувач системи, наділений наглядовими функціями; переглядає аналітику та результати обробки, контролює якість опрацювання вхідних запитів і створених Opportunity
Інтеграційна платформа (SAP CPI)	Системний актор, що забезпечує передачу JSON-представлення листа до AI Email Parser Agent, отримання структурованого результату та ініціювання створення/оновлення Opportunity у SAP C4C через API

Таблиця 3.9 - Виявлення варіантів використання акторів

Основний актор	Найменування	Формулювання
Адміністратор	Авторизація користувача	Дозволяє адміністратору входити в систему
Адміністратор	Моніторинг роботи системи	Дозволяє контролювати стан агентів
Менеджер з продажів	Перегляд результатів аналізу	Дозволяє отримувати структуровані дані
Менеджер з продажів	Формування звітів	Дозволяє формувати звіти

Керівник підрозділу	Перегляд аналітики	Дозволяє оцінювати ефективність
Інтеграційна платформа	Передача даних	Забезпечує обмін даними між системами

На рисунку 3.7 наведено діаграма варіантів використання

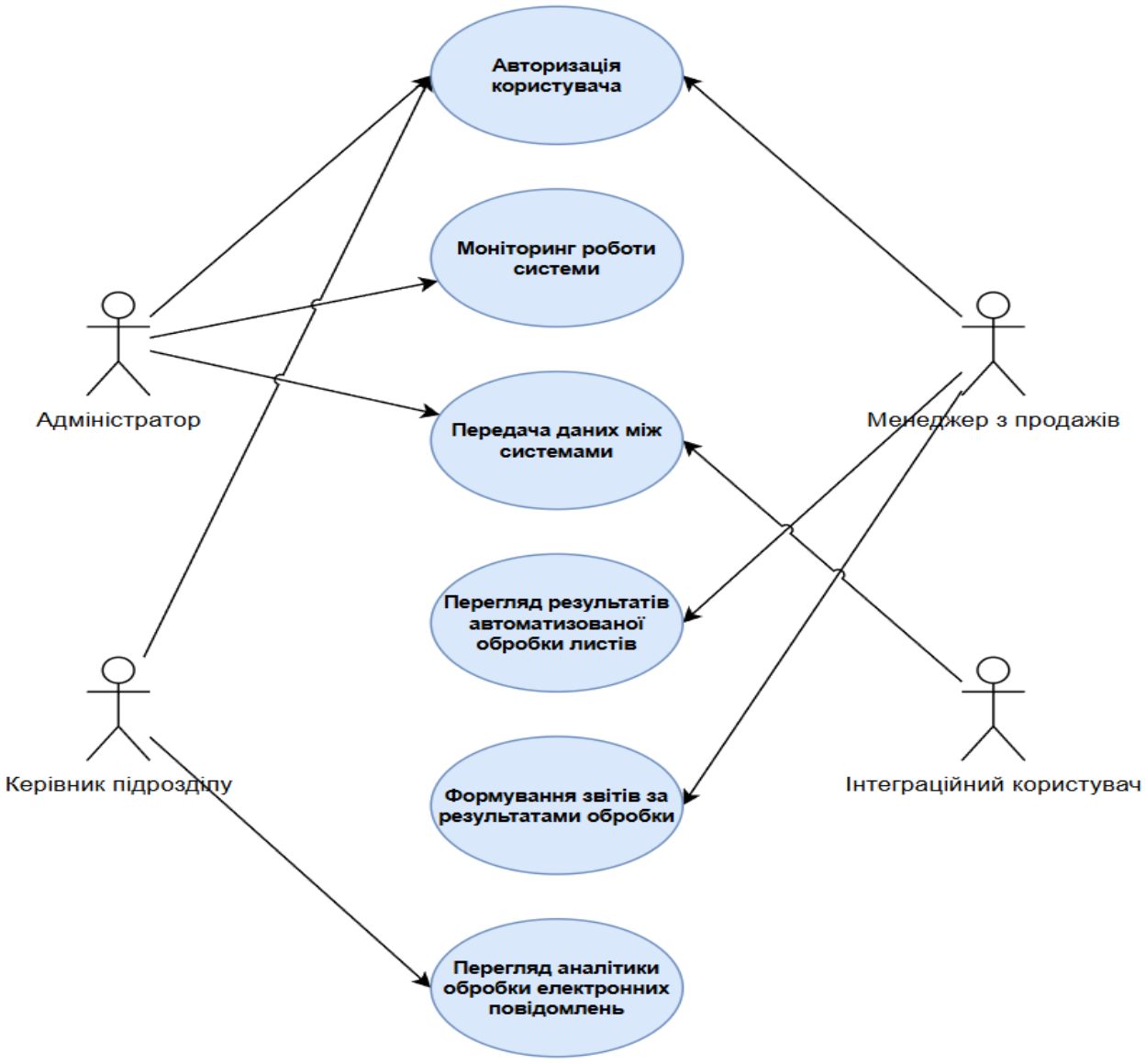

Рисунок 3.7 - Діаграма варіантів використання

На першому етапі була розроблена функціональна схема системи (рис. 3.8).

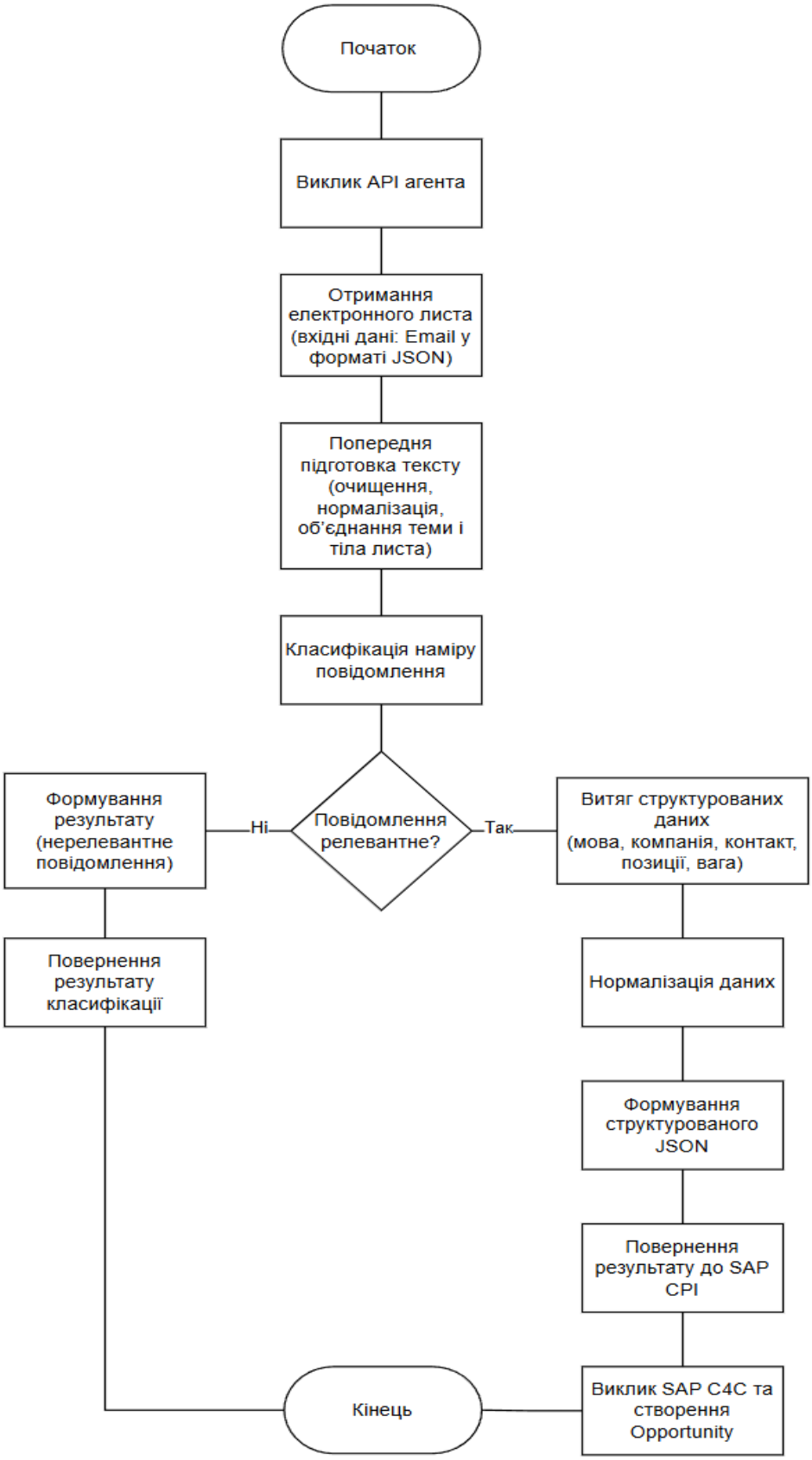

Рисунок 3.8 - Функціональна схема розроблюваної програмної системи

З наведеної функціональної схеми (рис. 3.8) випливає, що програмний комплекс AI Email Parser Agent доцільно декомпозувати на послідовність взаємопов’язаних функціональних блоків, кожен з яких відповідає окремому кроку перетворення неструктурованого листа на формалізований результат для CRM. Такий підхід узгоджується з принципами структурного проєктування ПЗ, де система розглядається як сукупність модулів із чітко визначеними входами/виходами та відповідальністю, що спрощує подальший супровід і масштабування.

Розглянемо ланцюг відповідає загальній логіці інтеграційно-орієнтованих систем, де шина інтеграції забезпечує оркестрацію, а прикладний сервіс - інтелектуальну обробку даних.

- Вхідне електронне повідомлення (Email/JSON)
- Інтеграційний компонент SAP CPI (приймання, трансформація, маршрутизація)
- Виклик API інтелектуального агента (/analyze-email)
- Класифікація наміру (релевантний/нерелевантний)
- Умова: якщо релевантний → перехід до витягу атрибутів; інакше → повернення мінімальної відповіді
- Витяг сутностей та нормалізація (мова, компанія, контакт, позиції, вага)
- Формування структурованого результату (JSON для CRM)
- Повернення відповіді в SAP CPI та мапінг полів під SAP C4C
- Виклик SAP C4C OData/REST та створення Opportunity (або підготовка запиту на створення).

У межах зазначеної функціональної схеми доцільно виокремити такі ключові блоки обробки: (1) приймання та приведення листа до єдиного формату (JSON) на рівні SAP CPI; (2) виклик аналітичного сервісу агента через REST API; (3) визначення релевантності

повідомлення методами машинного навчання (intent classification), що є типовою задачею автоматичної категоризації текстів [18]; (4) витяг сутностей (NER) та нормалізація атрибутів, що належить до базових напрямів обробки природної мови [18, 20]; (5) підготовка структурованого виходу, орієнтованого на подальшу обробку CRM-процесом у SAP C4C, де дані мають відповідати контракту інтеграції (поля, типи, обов'язковість).

На наступному етапі проєктування було сформовано ієрархічну схему модулів розроблюваної системи (рис. 3.9), яка відображає внутрішню декомпозицію програмного комплексу AI Email Parser Agent та логіку взаємодії між основними компонентами. Така модульна структуризація є типовим підходом у програмній інженерії, оскільки забезпечує керованість складності, спрощує супровід, тестування та дає можливість незалежно розвивати окремі підсистеми (наприклад, замінити модель класифікації або розширити блок витягу сутностей без зміни API-контракту).

На рис. 3.9 наведено ієрархію модулів, що складається з таких логічних блоків:

– A – базовий модуль приймання запиту та первинної валідації(перевірка структури вхідного JSON електронного листа, контроль обов'язкових полів, ініціалізація контексту обробки). У разі успішної валідації виконується передача керування до основного оркестратора процесу;

– Menu – головний модуль (оркестратор пайплайну), який визначає послідовність виконання етапів обробки: класифікація → витяг атрибутів → нормалізація → формування вихідного JSON. Також у цьому модулі реалізуються правила маршрутизації (наприклад, гілкування для нерелевантних листів) та централізоване логування ключових подій;

– AB_1 – модуль попередньої обробки повідомлення (pre-processing): нормалізація тексту, визначення мови, підготовка тексту/темы/підпису до аналізу, виділення сегментів (тіло листа, підпис, список позицій). За потреби виконується базове очищення та уніфікація формату даних;

– AB_2 – модуль Intent Classification, у якому виконується класифікація повідомлення на «релевантне/нерелевантне» (наприклад, запит комерційної пропозиції vs. службова/неробоча кореспонденція). Результат цього модуля визначає подальший сценарій: запуск витягу сутностей або завершення з короткою відповіддю;

– AB_3 – модуль витягу ключових атрибутів (NER + правила), що відповідає за визначення компанії, контактної особи, позицій номенклатури, кількості/ваги, одиниць виміру та інших доменних полів. На цьому етапі здійснюється також первинна перевірка узгодженості даних (наприклад, наявність ваги для позицій, формат чисел тощо);

– ABA_3 – допоміжний модуль постобробки та представлення результатів: нормалізація знайдених значень (уніфікація одиниць виміру, округлення, формування підсумкової ваги), генерація «short name» замовлення, а також приведення результату до узгодженого контракту відповіді (структурований JSON для SAP CPI / SAP C4C);

– AB_4 – модуль інтеграційного виходу (integration output), який формує фінальне повідомлення-відповідь для SAP CPI: заповнення полів згідно контракту, проставлення ознак релевантності, додавання службових метаданих (статус, помилки валідації, попередження), що дозволяє CPI коректно виконати мапінг та виклик SAP C4C OData/REST;

– AB_5 – модуль контролю якості та діагностики (quality control): перевірка коректності сформованої структури (schema validation), контроль повноти обов’язкових атрибутів, формування повідомлень про помилки/попередження та підготовка технічної інформації для логів і подальшого аналізу точності (у т.ч. для звітності під час експериментів).

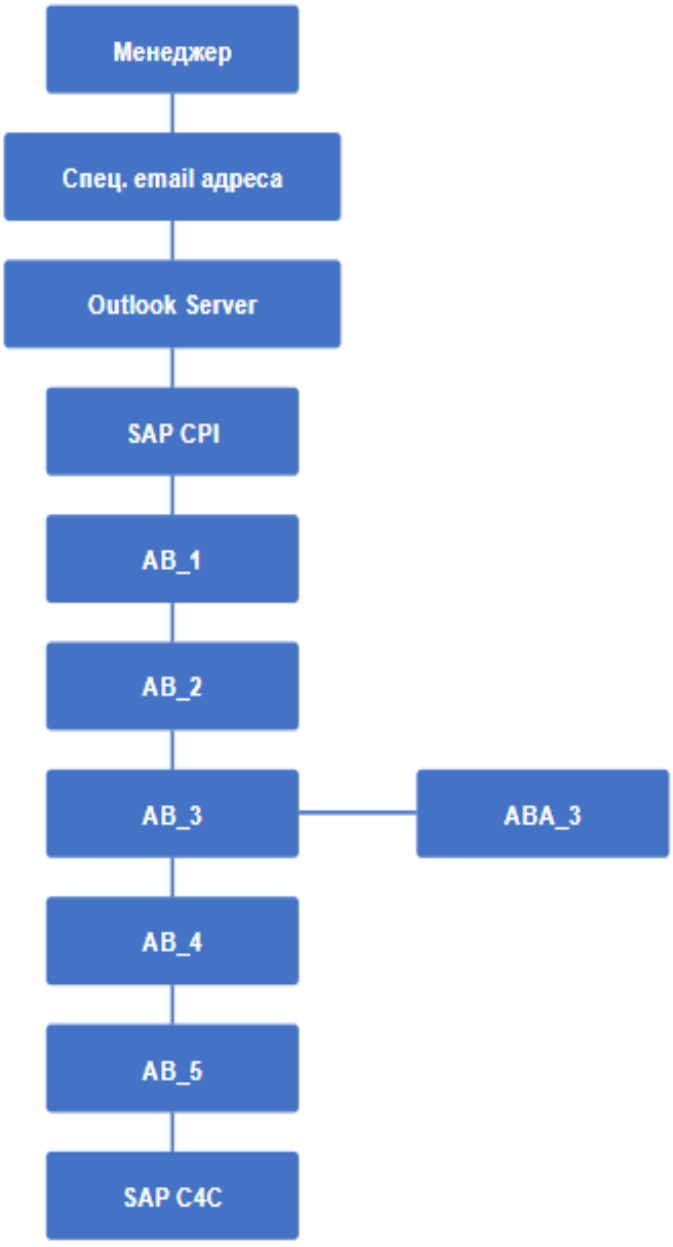

Рисунок 3.9 - Ієрархічна схема модулів розроблюваної системи

3.2 Розробка логічної та фізичної моделі програмного комплексу для автоматизації обробки вхідних електронних повідомлень у CRM-системі

Далі, з метою формалізації структури даних і взаємодії програмних компонентів, було розроблено UML-діаграму класів (рис. 3.10) для програмної системи AI Email Parser Agent. До складу діаграми доцільно включити такі основні класи:

- ApiRequest / EmailMessage – клас вхідного повідомлення (тема, тіло, відправник, отримувачі, час, вкладення/ознаки наявності);
- Preprocessor – підготовка тексту, виділення сегментів, визначення мови;
- IntentClassifier – класифікація «релевантне/нерелевантне» та повернення score/label;
- EntityExtractor (NERExtractor) – витяг сутностей (компанія, контакт, позиції, кількість/вага);
- Normalizer – нормалізація атрибутів та одиниць, підрахунок сумарної ваги, чистка значень;
- ShortNameBuilder – формування короткої назви замовлення на основі витягнутих позицій;
- AnalysisResult / ParsedOrder – клас результату аналізу (релевантність, мова, компанія, контакт, список позицій, total_weight, short_name, повідомлення);
- ResponseBuilder – складання фінального JSON-відповіді за інтеграційним контрактом для SAP CPI/SAP C4C;
- QualityGate / Validator – перевірка схеми, обов’язкових полів та формування попереджень/помилок.

Такий набір класів забезпечує чіткий розподіл відповідальності між шарами системи (вхід/обробка/вихід/контроль якості) та відповідає підходу separation of concerns, який є базовим для проєктування підтримуваних програмних рішень.

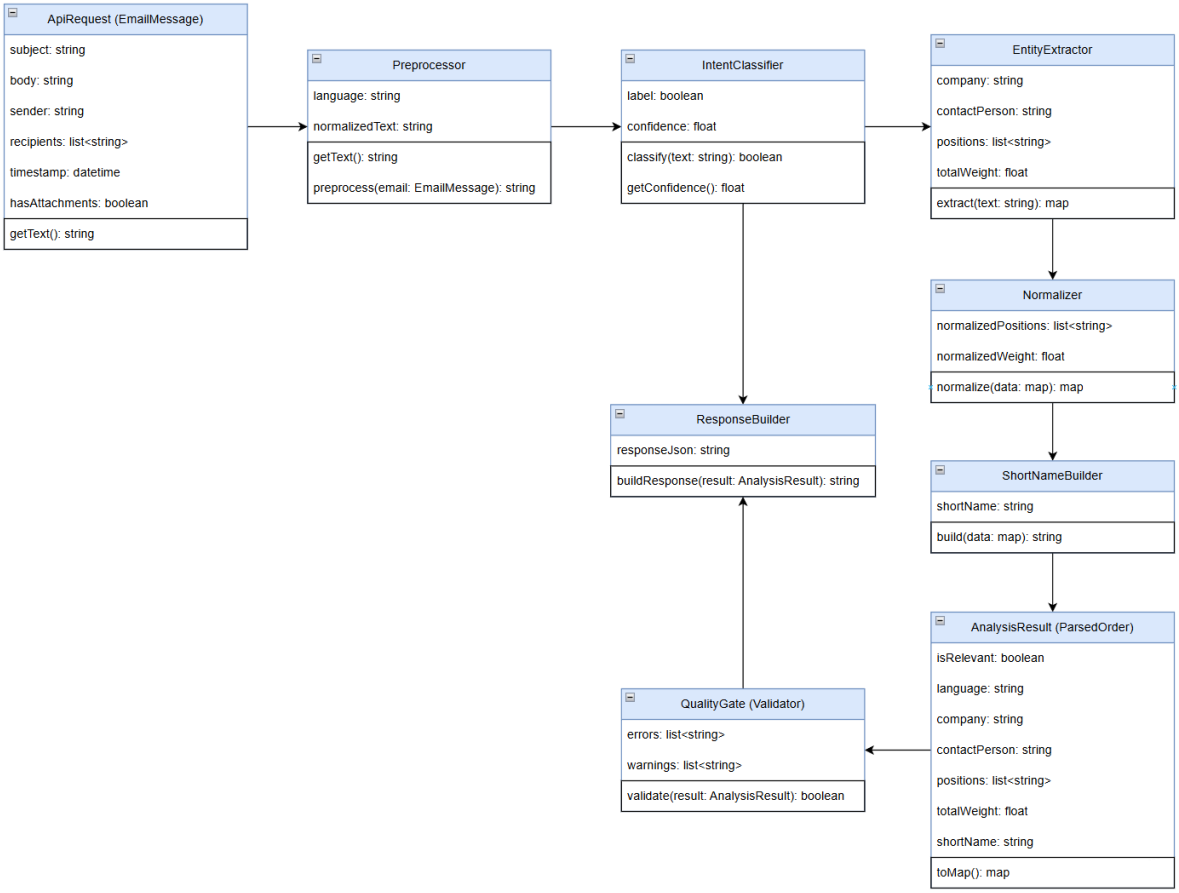

Рисунок 3.10 - Діаграма класів UML для програмної системи, що розробляється

Як було зазначено у попередніх розділах, у сучасних умовах цифрової трансформації підприємств однією з найбільш актуальних проблем є ефективна обробка та інтерпретація неструктурованої текстової інформації, зокрема вхідної електронної кореспонденції. Обсяг електронних листів, що щоденно надходять до корпоративних

поштових скриньок від клієнтів, партнерів і контрагентів, постійно зростає, що призводить до суттєвого навантаження на персонал та збільшення часу реакції на запити [29, 31].

У більшості організацій обробка вхідних листів, які містять комерційні запити, замовлення або технічні звернення, виконується вручну. Це включає аналіз тексту листа, визначення його релевантності, ідентифікацію компанії та контактної особи, виділення переліку позицій, кількісних характеристик і подальше перенесення цих даних у CRM-системи. Такий підхід є трудомістким, залежним від людського фактора та схильним до помилок, що особливо критично для великих підприємств із високим потоком звернень, про що свідчать галузеві аналітичні дослідження [29, 30].

З економічної точки зору ручна обробка текстових звернень призводить до збільшення операційних витрат, втрати потенційних клієнтів через затримки в обробці запитів, а також до зниження якості даних у корпоративних інформаційних системах. Крім того, відсутність єдиного формалізованого підходу до обробки листів ускладнює подальшу аналітику, прогнозування та автоматизацію бізнес-процесів, що суперечить принципам управління якістю та ефективності інформаційних систем [18, 19].

Одним із перспективних шляхів вирішення зазначених проблем є використання інтелектуальних агентів штучного інтелекту, здатних автоматично аналізувати текст електронних листів, виконувати їх класифікацію та витяг ключових атрибутів з подальшим формуванням структурованих даних. Застосування методів обробки природної мови, машинного навчання та агентних підходів дозволяє значно підвищити швидкість обробки звернень, зменшити кількість помилок та забезпечити стабільну якість даних, що передаються до CRM-систем [1, 3, 11].

Особливої актуальності набуває інтеграція таких інтелектуальних рішень із корпоративними платформами класу SAP Cloud for Customer та SAP Cloud Platform Integration, що широко використовуються на великих промислових підприємствах. Автоматизоване створення бізнес-об’єктів, зокрема Opportunity, на основі результатів аналізу електронних листів дозволяє скоротити час обробки запитів, підвищити прозорість процесів продажу та забезпечити більш ефективну взаємодію між підрозділами відповідно до сучасних підходів інтеграції корпоративних систем [17, 32].

Таким чином, дослідження моделей, методів та інформаційних технологій використання агентів штучного інтелекту для автоматизації обробки електронних листів є актуальним науково-практичним завданням, що відповідає сучасним тенденціям розвитку корпоративних інформаційних систем і спрямоване на підвищення ефективності операційної діяльності підприємств [29, 31].

У межах дослідження теми «Дослідження предметної області та методів використання агентів штучного інтелекту для автоматизації обробки даних» було виконано аналіз предметної області, пов’язаної з автоматизованою обробкою вхідної електронної кореспонденції у корпоративних CRM-системах. Результатом проведеного аналізу стала формалізація ключових взаємодій користувачів та зовнішніх інформаційних компонентів із програмним комплексом AI Email Parser Agent for SAP C4C Opportunity, що дало змогу побудувати діаграму варіантів використання (Use Case) для розроблюваної системи (рис. 3.11).

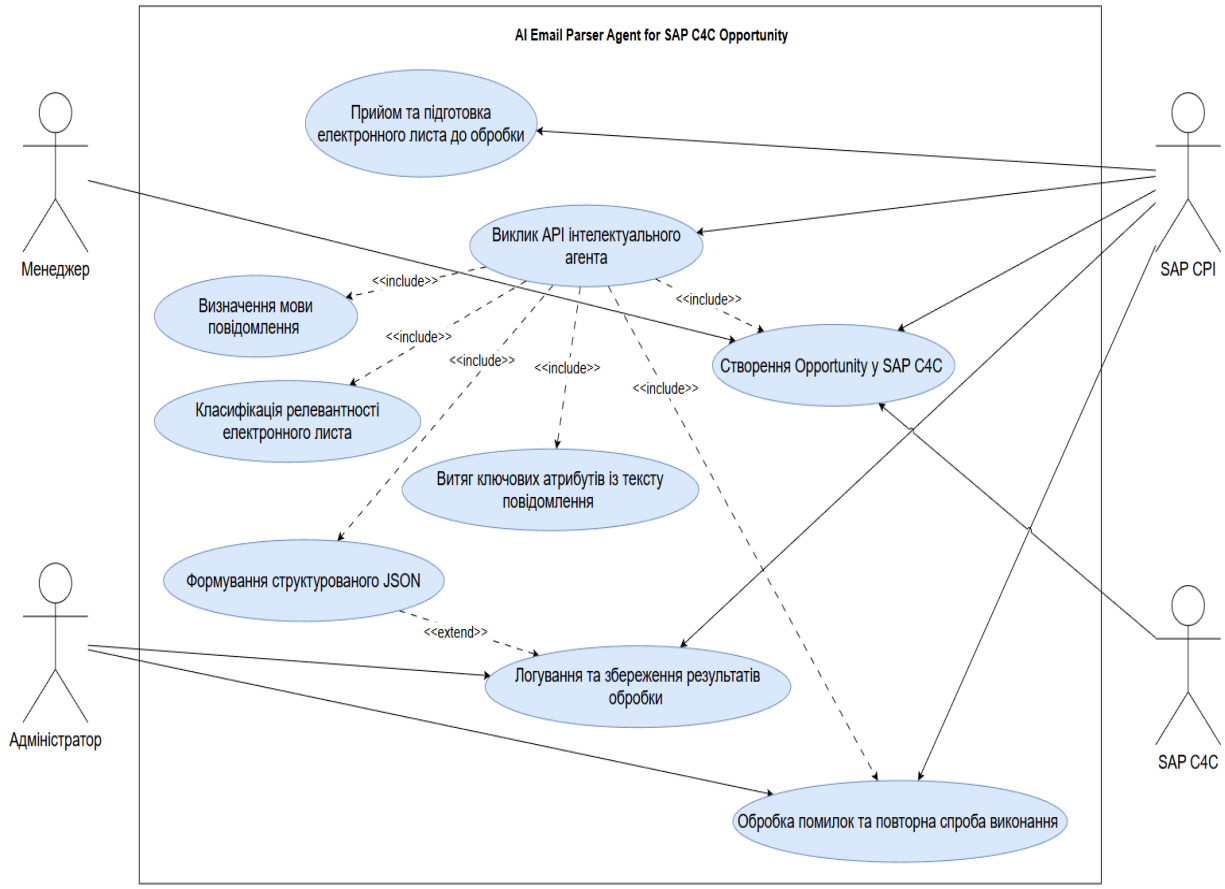

Рисунок 3.11 – Діаграма Use Case для програмної системи, що розробляється

У розроблюваній системі доцільно виділити такі основні актори (користувачі та зовнішні системи), які взаємодіють із програмним комплексом:

– менеджер з продажів / користувач CRM – ініціює та використовує бізнес-процес обробки вхідних звернень клієнтів, очікуючи автоматичного перетворення неструктурованих електронних листів у структуровані дані та створення комерційних можливостей (Opportunity) у системі SAP C4C;

– адміністратор системи – відповідає за працездатність сервісу, виконує налаштування інтеграційних сценаріїв і правил маршрутизації, здійснює контроль журналів подій та діагностику помилок;

– поштова система (email-сервер або поштова скринька) – зовнішнє джерело вхідних електронних повідомлень, яке ініціює подію обробки листа;

– SAP CPI (інтеграційний шар) – забезпечує оркестрацію інтеграційного сценарію, включаючи прийом електронного листа, підготовку та трансформацію запиту, виклик API інтелектуального агента, обробку відповіді та інтеграцію з CRM-системою SAP C4C;

– AI Email Parser Agent (серверний API-сервіс) – реалізує інтелектуальну обробку тексту електронного повідомлення, виконує класифікацію релевантності та витяг структурованих атрибутів із подальшим формуванням JSON-результату;

– SAP C4C (CRM-система) – приймає результати обробки та забезпечує створення або оновлення бізнес-об’єктів типу Opportunity разом із пов’язаними атрибутами.

Для зазначених акторів визначено такі основні варіанти використання програмної системи:

– прийом та підготовка електронного листа до обробки (актор: SAP CPI; джерело події: поштова система). Варіант використання передбачає нормалізацію вхідних даних (тема, тіло листа, підпис, текстові фрагменти та вкладення) і формування уніфікованого JSON-запиту для подальшого аналізу.

– виклик API інтелектуального агента (/analyze-email) (актор: SAP CPI → AI Email Parser Agent). Є ключовою інтеграційною операцією, яка ініціює процес інтелектуального аналізу електронного повідомлення.

– визначення мови повідомлення (*include*, актор: AI Email Parser Agent). Забезпечує вибір коректної стратегії обробки тексту та застосування відповідних правил нормалізації.

– класифікація релевантності електронного листа (*include*, актор: AI Email Parser Agent). Результатом є визначення статусу повідомлення (релевантне або нерелевантне) щодо доцільності створення Opportunity у CRM-системі.

– витяг ключових атрибутів із тексту повідомлення (*include*, актор: AI Email Parser Agent). Передбачає визначення назви компанії, контактної особи, переліку товарних позицій, кількості або ваги замовлення, а також формування короткої назви (short name) звернення.

– формування структурованого JSON-висновку (*include*, актор: AI Email Parser Agent). Сформований результат повертається до SAP CPI та використовується для подальших етапів автоматизації бізнес-процесу.

– створення Opportunity у SAP C4C (*include*, актор: SAP CPI → SAP C4C). Виконується виключно для релевантних повідомлень та включає мапінг атрибутів і ініціацію OData/REST-виклику.

– логування та збереження результатів обробки (актор: SAP CPI / AI Email Parser Agent; адміністрування: адміністратор системи). Забезпечує аудит дій, трасування помилок та можливість повторної обробки повідомлень.

– обробка помилок та повторна спроба виконання (*extend*, актор: SAP CPI; контроль: адміністратор системи). Активується у випадках недоступності сервісу інтелектуального агента, помилок валідації даних або відмови CRM-системи SAP C4C.

На наступному етапі проєктування була сформована діаграма діяльності програмної системи, що розробляється (рис. 3.12), яка відображає логіку виконання основних бізнес-процесів та послідовність дій користувача і системи під час обробки вхідних електронних листів.

Початковим етапом функціонування системи є надходження нового електронного листа або ініціювання потреби в його аналізі. Далі користувач (у даному випадку технолог або відповідальний менеджер) проходить процедуру авторизації в системі. Після цього відбувається перевірка коректності облікових даних. У разі успішної авторизації система переходить до наступного етапу - отримання та підготовки початкових даних для аналізу, а саме: тексту листа, теми повідомлення, метаданих відправника та вкладень (за наявності).

На основі підготовлених даних виконується початковий аналіз електронного листа із застосуванням інтелектуального агента. Якщо отримані результати (класифікація релевантності та витяг ключових атрибутів) є задовільними, система переходить до наступних етапів обробки. У разі отримання некоректних або неповних результатів ініціюється додатковий аналіз, що може включати повторний запуск моделі або використання альтернативних правил обробки.

Після завершення аналізу результати структуруються, зберігаються у базі даних, формуються у вигляді звіту та передаються до зовнішніх інформаційних систем, зокрема до CRM-системи SAP C4C для автоматичного створення або оновлення об’єкта Opportunity. Надалі здійснюється контроль коректності переданих даних та відповідності створених бізнес-об’єктів результатам аналізу електронного листа.

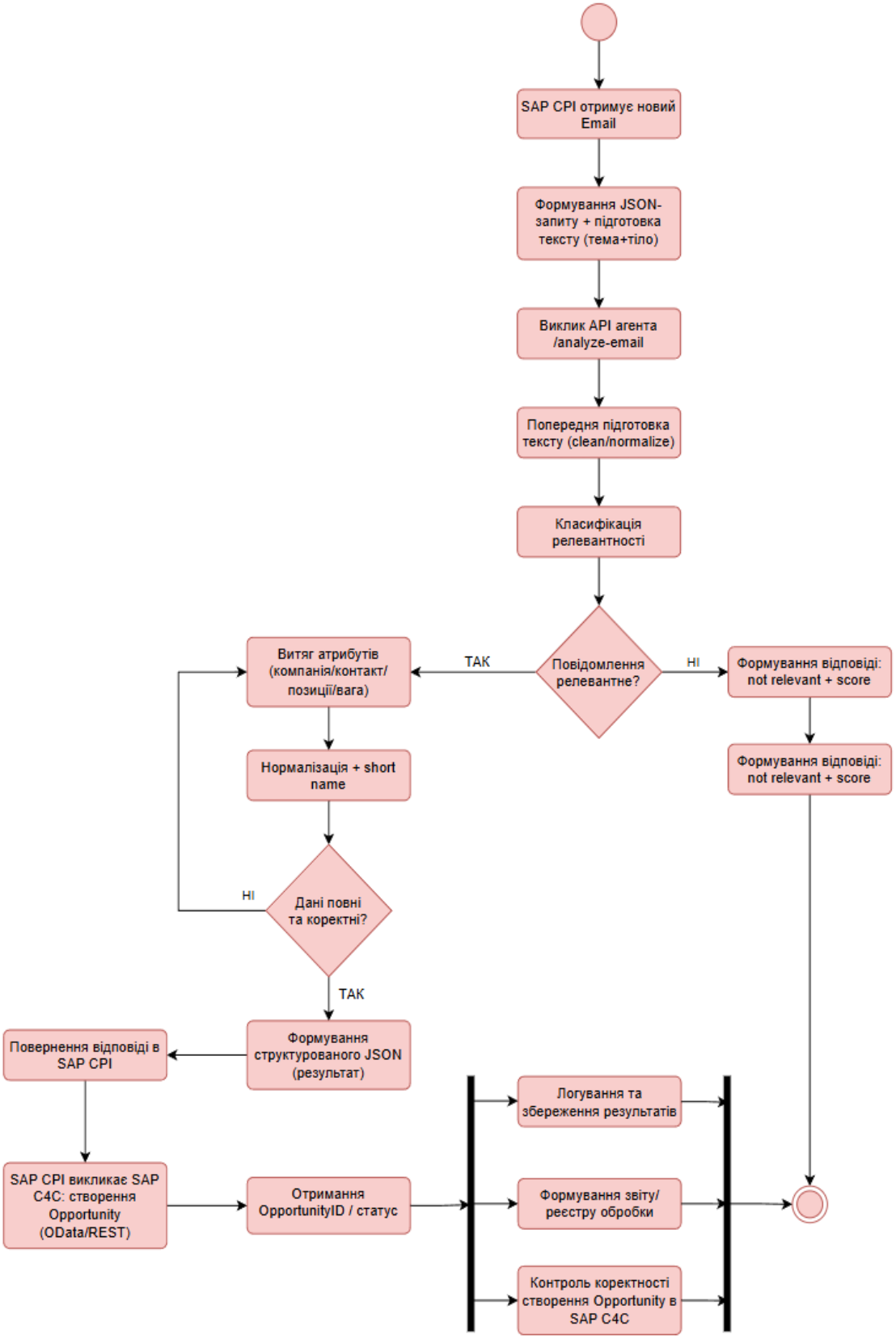

Рисунок 3.12 - Діаграма діяльності програмної системи, що розробляється

Для формалізації поведінки системи у процесі обробки електронних листів була розроблена діаграма станів (State) програмної системи, що розробляється (рис. 3.13).

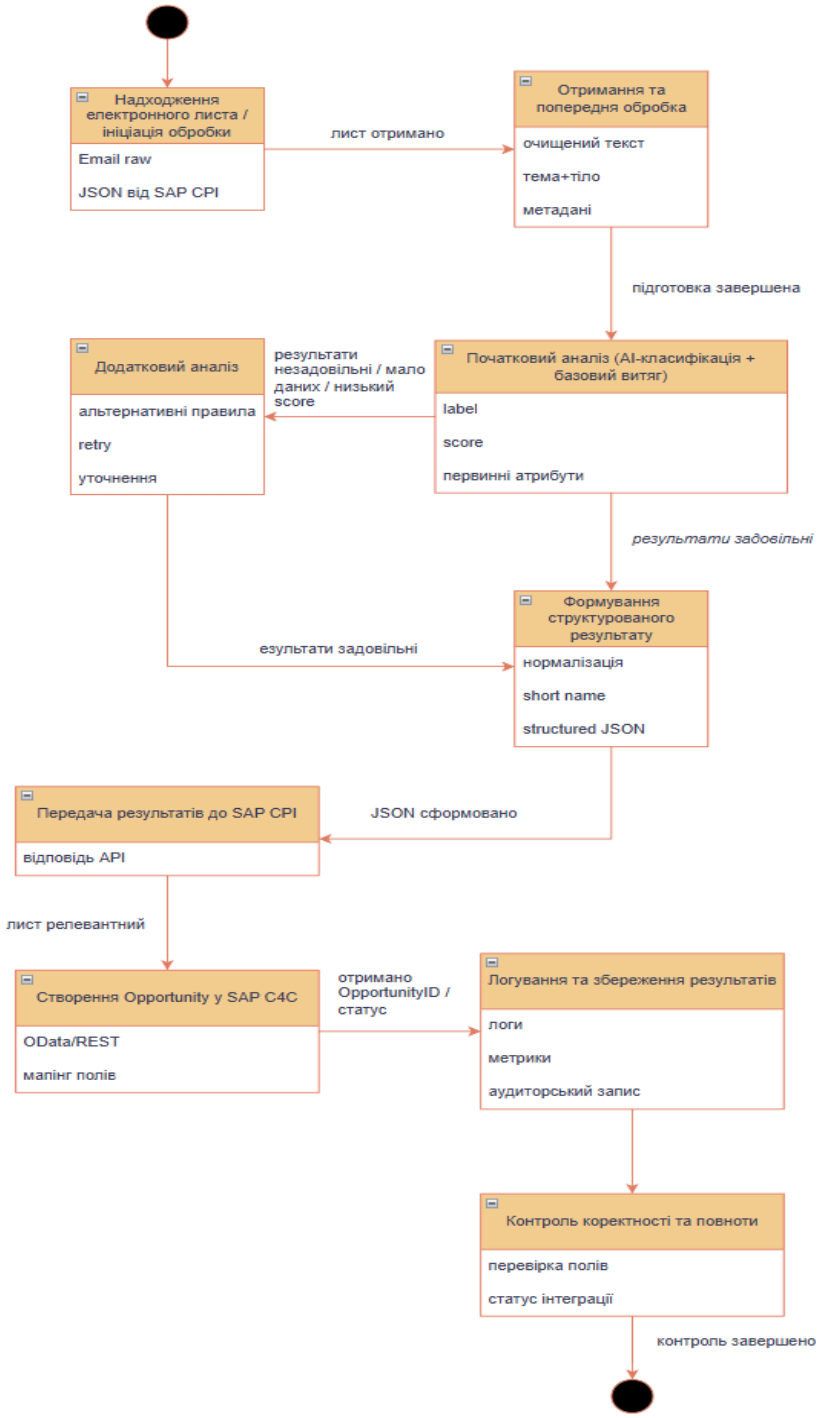

Рисунок 3.13 - Діаграма станів (State) для програмної системи, що розробляється

Також була побудована продукційна модель для програмної системи, що розробляється, яка формалізує логіку прийняття рішень та послідовність виконання дій у вигляді правил типу «якщо – то». Застосування продукційної моделі дозволяє описати поведінку системи з точки зору переходів між станами на основі наявних фактів та виконуваних дій, що є доцільним для систем інтелектуальної обробки даних.

Цільовими (обов’язковими) діями при функціонуванні програмної системи є:

– аналіз вхідного електронного листа із використанням інтелектуального агента;

– передача структурованих результатів обробки до CRM-системи SAP C4C для подальшого створення або оновлення об’єкта Opportunity.

Проміжними діями при функціонуванні системи є:

– авторизація користувача або сервісного облікового запису інтеграційної платформи;

– отримання та попередня обробка електронного листа;

– виконання початкового аналізу електронного листа;

– виконання додаткового аналізу у разі отримання незадовільних або неповних результатів;

– формування структурованих результатів аналізу;

– формування звіту або реєстру обробки;

– збереження результатів аналізу у базі даних;

– передавання результатів до CRM-системи SAP C4C;

– контроль коректності та повноти переданих даних.

У разі надходження електронного листа, що потребує обробки, ініціюється процес авторизації користувача або сервісного облікового запису, від імені якого виконується інтеграційний сценарій. Після

успішної авторизації система отримує початкові дані електронного листа, зокрема текст повідомлення, тему, метадані відправника та вкладення (за наявності), і виконує їх попередню підготовку до аналізу.

На наступному етапі запускається початковий аналіз електронного листа із застосуванням інтелектуального агента, який виконує класифікацію релевантності повідомлення та витяг ключових атрибутів. Якщо результати початкового аналізу є незадовільними, наприклад у разі низької впевненості класифікації або неповного витягу даних, система ініціює додатковий аналіз, що може включати повторний запуск моделі або використання альтернативних правил обробки.

Після отримання коректних та узгоджених результатів аналізу система формує структурований результат, який містить усі необхідні атрибути для подальшої автоматизації бізнес-процесу. Сформовані результати зберігаються у базі даних, а також передаються до CRM-системи SAP C4C для створення або оновлення об’єкта Opportunity. На завершальному етапі здійснюється контроль коректності та повноти переданих даних з метою забезпечення цілісності бізнес-інформації.

Узагальнюючи описану логіку, процес функціонування програмної системи може бути поданий у вигляді наступних правил типу «якщо – то»:

– якщо надійшов новий електронний лист, то виконується авторизація користувача або сервісного облікового запису;

– якщо авторизація пройшла успішно, то здійснюється отримання та підготовка початкових даних електронного листа;

– якщо початкові дані підготовлено, то запускається початковий аналіз електронного листа;

– якщо результати початкового аналізу є незадовільними, то виконується додатковий аналіз;

– якщо результати аналізу є задовільними, то формуються структуровані результати обробки;

– якщо сформовано результати аналізу, то вони зберігаються у базі даних;

– якщо сформовано результати аналізу, то вони передаються до CRM-системи SAP C4C;

– якщо результати передано до CRM-системи, то виконується контроль коректності переданих даних.

Для формалізації продукційної моделі програмної системи введемо позначення фактів (Ф), дій (Д) та продукцій (П).

Суб’єктом продукційної моделі є вхідний електронний лист, що надходить до системи для подальшої інтелектуальної обробки.

Позначення фактів:

Ф1 – отримано електронний лист, що потребує аналізу;

Ф2 – необхідно отримати структуровані результати аналізу електронного листа;

Ф3 – результати аналізу електронного листа сформовано;

Ф4 – результати аналізу передано до CRM-системи та підлягають перевірці коректності.

Позначення дій:

Д1 – користувач або сервісний обліковий запис успішно авторизувався у системі;

Д2 – отримано та підготовлено початкові дані електронного листа для аналізу;

Д3 – виконується початковий аналіз електронного листа із застосуванням інтелектуального агента;

Д4 – виконується додатковий аналіз електронного листа з метою уточнення результатів;

Д5 – сформовано структуровані результати аналізу та звіт за результатами обробки;

Д6 – результати аналізу збережено у базі даних;

Д7 – результати аналізу передано до CRM-системи SAP C4C;

Д8 – виконується контроль коректності та повноти переданих даних.

Зробимо опис продукційних правил.

Для кожної продукції встановлено пріоритет (у дужках перед комою), при цьому чим більший пріоритет, тим раніше перевіряється відповідне правило:

П1 (8, Ф1) → Д1;

П2 (7, Ф1 та Д1) → Д2;

П3 (6, Ф2 та Д2) → Д3;

П4 (5, Ф2 та Д3) → Д4;

П5 (4, (Ф3 та Д3) або (Ф3 та Д4)) → Д5;

П6 (3, Ф3 та Д5) → Д6;

П7 (2, Ф3 та Д6) → Д7;

П8 (1, Ф4 та Д7) → Д8.

Таким чином, продукційна модель описує послідовний процес переходу системи від факту надходження електронного листа до етапу контролю коректності переданих результатів у CRM-системі. Модель враховує можливість виконання додаткового аналізу у разі отримання незадовільних результатів первинної обробки, що підвищує надійність та якість функціонування програмної системи.

Для наочного відображення взаємозв’язків між продукціями та відповідними діями побудовано граф продукції, наведений на рисунку 3.14.

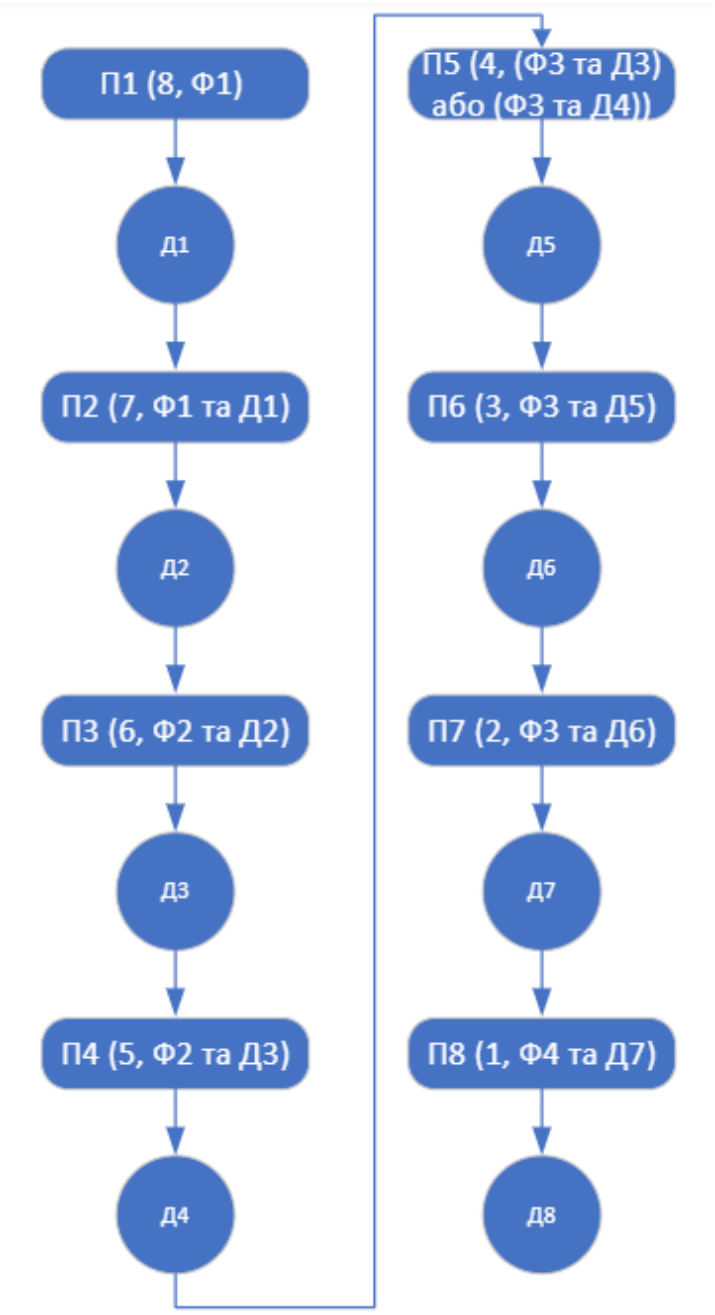

Рисунок 3.14 – Схема продукції предметної області програмної системи AI Email Parser Agent.

Для формалізації знань у предметній області автоматизованої обробки вхідних електронних листів була побудована мережева (семантична) модель представлення знань. Дана модель відображає

ключові сутності предметної області, їх взаємозв’язки та логіку функціонування програмної системи - інтелектуального агента аналізу електронної кореспонденції.

Метою побудови семантичної мережі є формальний опис процесів класифікації, аналізу та перетворення неструктурованих текстових повідомлень у структуровані дані, придатні для подальшої автоматизації бізнес-процесів у корпоративних інформаційних системах.

Ключовими поняттями предметної області є: електронний лист, користувач системи, інтеграційна платформа, інтелектуальний агент аналізу, початкові дані листа, результати обробки, структуровані дані, база даних та зовнішні інформаційні системи (CRM).

Вершинами семантичної мережі визначено такі сутності: користувач, електронний лист, AI Email Parser Agent, результати аналізу, звіт, база даних, CRM-система (SAP C4C) та бізнес-об’єкт Opportunity.

Логіка взаємодії полягає в тому, що користувач або інтеграційна підсистема ініціює процес обробки вхідного електронного листа. Інтелектуальний агент виконує аналіз текстового вмісту повідомлення, здійснює класифікацію релевантності, витягує ключові атрибути та формує структурований результат. Отримані результати використовуються для формування звіту, збереження в базі даних та передачі до зовнішньої CRM-системи для створення або оновлення бізнес-об’єктів.

Семантична мережа предметної області наведена на рисунку 3.15.

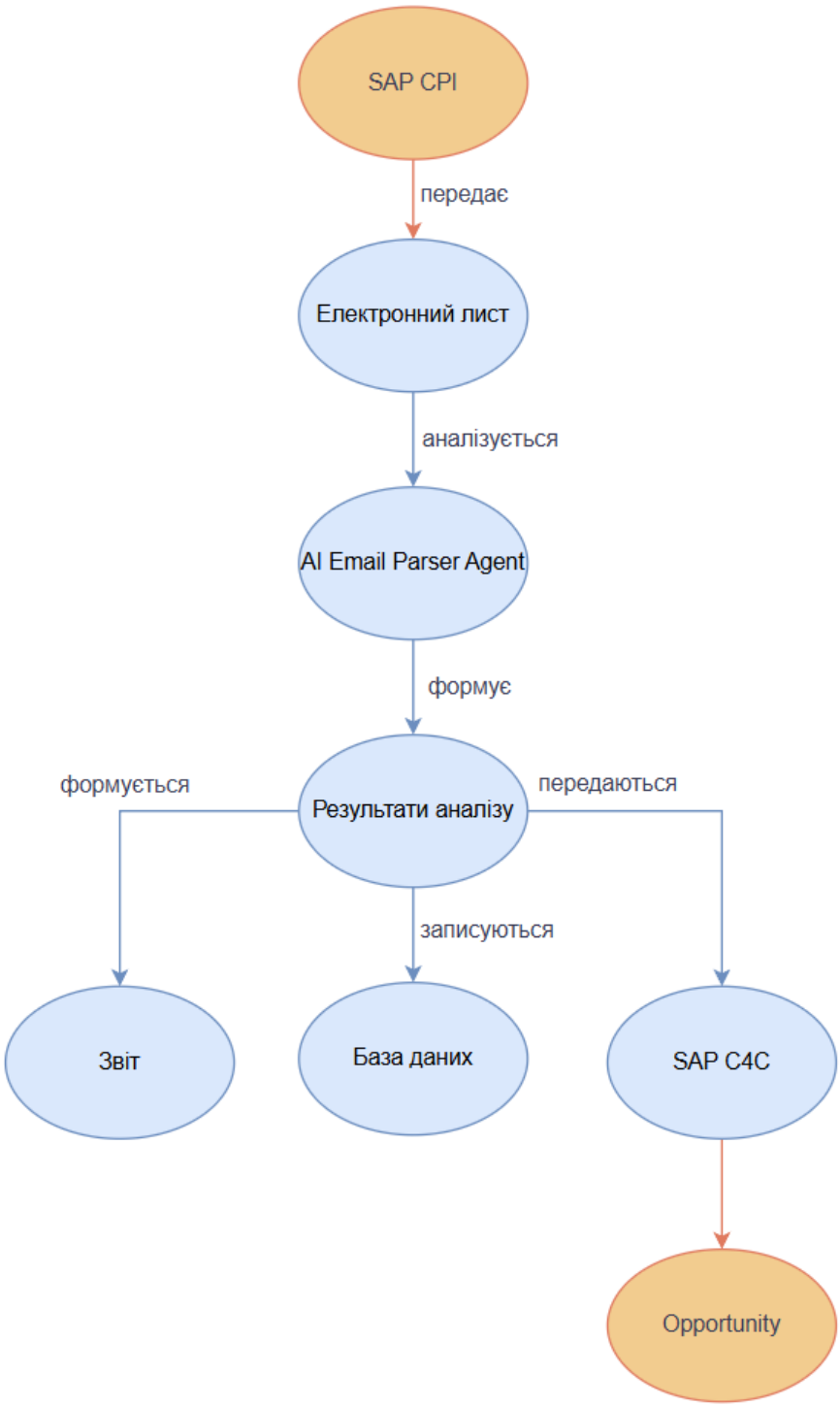

Рисунок 3.15 - Семантична мережа предметної області програмної системи, що розробляється

Для більш детального опису об’єктів предметної області побудовано фреймову модель подання знань. Основними фреймами є користувач, електронний лист та програмне забезпечення (AI-агент) (табл. 3.10). Кожен фрейм описується набором слотів, що визначають його атрибути та характеристики.

Таблиця 3.10 - Опис фреймів – об’єктів

Користувач			
Ім’я слоту	Значення слоту	Спосіб отримання значення	Демон
Роль	Керуючий / базовий	З зовнішніх джерел	–
Посада	Адміністратор / Бізнес-користувач / Аналітик	З зовнішніх джерел	–
Електронний лист			
Ім’я слоту	Значення слоту	Спосіб отримання значення	Демон
Джерело	Корпоративна або зовнішня пошта	Автоматично	–
Мова	Українська / Англійська / Інша	NLP-аналіз	–
Тип	Комерційний / Інформаційний / Нерелевантний	Класифікація	–
Програмне забезпечення			
Ім’я слоту	Значення слоту	Спосіб отримання значення	Демон
Призначення	Аналіз та структуризація листів	Зовнішні джерела	–
Рівень	Корпоративний / Інтеграційний	Зовнішні джерела	–

Фрейми-спадкоємці успадковують усі слоти базового фрейму «Користувач» і деталізують їх залежно від ролі в системі. Опис фреймів-спадкоємців наведено у таблиці 3.12.

Таблиця 3.11 – Опис фреймів-зразків

Електронний лист (АКО Інформаційний об’єкт)			
Ім'я слоту	Значення слоту	Спосіб отримання значення	Демон
Призначення	Запит / Комерційна пропозиція / Повідомлення	NLP-аналіз	
Структура	Неструктурований текст	Вхідні дані	

Таблиця 3.12 - Опис фреймів-спадкоємців для об’єкту «Користувач»

Адміністратор (АКО Користувач)			
Ім'я слоту	Значення слоту	Спосіб отримання значення	Демон
Роль	Адміністратор	З зовнішніх джерел	
Посада	Адміністратор системи	З зовнішніх джерел	
Бізнес-користувач			
Ім'я слоту	Значення слоту	Спосіб отримання значення	Демон
Роль	Користувач	З зовнішніх джерел	
Посада	Менеджер з продажу / Аналітик	З зовнішніх джерел	

Для формалізації логіки функціонування програмної системи побудовано продукційну модель. Цільовими діями системи є аналіз вхідного електронного листа та формування структурованих даних для подальшої автоматизації бізнес-процесів.

Проміжними діями є:

- авторизація користувача або системи-ініціатора;
- отримання вхідного електронного листа;
- виконання початкового аналізу тексту;
- класифікація листа;
- витяг ключових атрибутів;
- формування структурованого результату;
- збереження результатів у базі даних;
- передавання даних до зовнішніх корпоративних систем.

Логіку роботи системи можна подати у вигляді правил типу «якщо – то»:

- якщо отримано новий електронний лист, то ініціюється процес аналізу;
- якщо аналіз виконано успішно, то формується структурований результат;
- якщо результат сформовано, то дані зберігаються у базі даних;
- якщо налаштовано інтеграцію, то результати передаються у CRM-систему.

Для відображення взаємозв’язків між продукціями побудовано граф продукцій, наведений на рисунку 3.16.

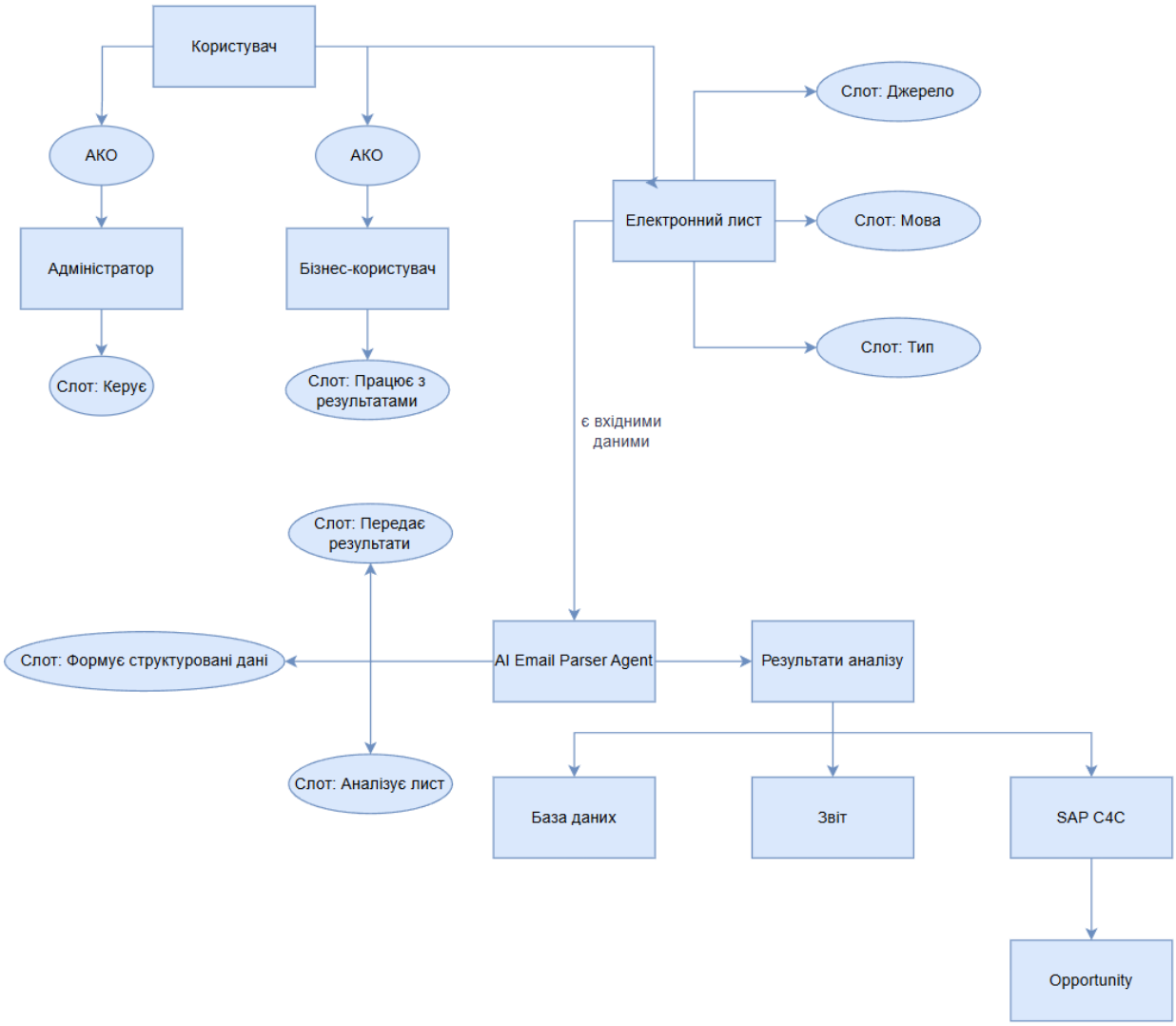

Рисунок 3.16 - Взаємозв’язок фреймів

3.3 Програмна реалізація інтелектуального агента

Програмна реалізація інтелектуального агента AI Email Parser Agent виконувалася з урахуванням вимог до побудови програмно-методичних комплексів, орієнтованих на дослідження та експериментальну перевірку моделей і методів штучного інтелекту.

Основною метою реалізації на даному етапі є створення працездатного прототипу (Proof of Concept), який дозволяє відтворити повний цикл обробки вхідного електронного повідомлення, перевірити коректність обраних алгоритмічних рішень та оцінити їх ефективність у контрольованих умовах. Відповідно до рекомендацій з проєктування програмних систем дослідницького характеру, реалізація орієнтована не на оптимізацію продуктивності, а на прозорість архітектури, відтворюваність експериментів і можливість подальшого розширення функціональності [1], [34].

При розробці інтелектуального агента було обрано модульний підхід до організації програмного коду, за якого окремі функціональні компоненти реалізуються у вигляді самостійних модулів із чітко визначеними інтерфейсами взаємодії. Такий підхід дозволяє ізолювати реалізацію моделей машинного навчання, алгоритмів обробки природної мови та логіки оркестрації від сервісного рівня доступу до агента. Модульна структура спрощує тестування окремих компонентів, полегшує аналіз результатів експериментів і відповідає принципам супроводжуваності та повторного використання програмного забезпечення, що є важливими для систем цифрового інтелекту в корпоративному середовищі [3], [5].

3.3.1 Структура проєкту та опис основних модулів

Структура програмного проєкту AI Email Parser Agent сформована відповідно до принципів модульності та логічного розподілу відповідальності між компонентами системи, що є типовим підходом при реалізації інтелектуальних сервісів дослідницького

характеру. Така організація програмного коду дозволяє чітко відокремити рівень сервісного доступу від рівня алгоритмічної обробки даних, а також забезпечує зручність тестування, аналізу результатів і подальшої модифікації окремих модулів без необхідності суттєвих змін у всій системі. Застосування подібного підходу відповідає рекомендаціям програмної інженерії щодо побудови масштабованих і супроводжуваних систем, що використовують методи машинного навчання та обробки природної мови [3], [34].

У межах реалізованого прототипу програмний комплекс організовано як набір функціонально завершених модулів, кожен з яких відповідає за окремий етап обробки електронного повідомлення. Центральним елементом є модуль-оркестратор, який керує послідовністю викликів моделей класифікації та витягу сутностей, забезпечуючи реалізацію ієрархічного алгоритму обробки. Окремо виділено модулі для навчання та застосування моделі класифікації наміру повідомлення, модуль аналізу тексту і витягу іменованих сутностей, а також допоміжні компоненти для генерації та підготовки датасету і формування звітності за результатами експериментів. Такий розподіл дозволяє розглядати кожен модуль як самостійний об’єкт дослідження та спрощує аналіз впливу окремих алгоритмічних рішень на загальну ефективність системи [1], [5].

Для забезпечення доступу до функціональності інтелектуального агента використовується окремий сервісний модуль, який реалізує REST API та відповідає за прийом вхідних даних, їх первинну валідацію і передачу на обробку внутрішнім компонентам системи. Винесення API-рівня в окремий модуль дозволяє відокремити логіку взаємодії із зовнішніми системами від внутрішньої реалізації алгоритмів, що є важливим з точки зору інтеграції з корпоративними платформами та дотримання принципів сервісно-

орієнтованої архітектури [35]. Крім того, така структура створює основу для подальшого масштабування рішення та можливого розгортання інтелектуального агента як незалежного сервісу в корпоративному інтеграційному середовищі.

3.3.2 Реалізація API для аналізу тексту (FastAPI)

Для забезпечення взаємодії інтелектуального агента з зовнішніми інформаційними системами програмний комплекс AI Email Parser Agent реалізовано у вигляді сервісу з REST-орієнтованим прикладним програмним інтерфейсом. Як технологічну основу для побудови API обрано фреймворк FastAPI, який широко застосовується для створення високопродуктивних сервісів обробки даних та машинного навчання. Використання REST API дозволяє стандартизувати обмін інформацією між інтелектуальним агентом та інтеграційною платформою SAP Cloud Platform Integration, а також забезпечує незалежність сервісу від конкретних реалізацій споживачів даних. Такий підхід відповідає сучасним принципам сервісно-орієнтованої архітектури та рекомендований для інтеграції інтелектуальних компонентів у корпоративні інформаційні системи [34], [35].

API інтелектуального агента реалізує єдину точку входу для аналізу електронних повідомлень, яка приймає вхідні дані у форматі JSON та повертає структурований результат обробки. На рівні API здійснюється первинна валідація вхідних параметрів, контроль цілісності даних та передача повідомлення до модуля-оркестратора для подальшого аналізу. Така організація дозволяє відокремити

задачі перевірки коректності запиту від алгоритмічної логіки обробки тексту, що підвищує надійність сервісу та спрощує його супроводження. Формат відповіді API уніфікований і містить як результати класифікації релевантності, так і витягнуті та нормалізовані атрибути, що робить його безпосередньо придатним для використання у сценаріях автоматизованого створення бізнес-об’єктів у CRM-системах [1], [3].

3.3.3 Технологічний стек програмного комплексу

Під час реалізації програмного комплексу AI Email Parser Agent було сформовано технологічний стек, орієнтований на розробку інтелектуальних сервісів аналізу текстових даних та їх інтеграцію в корпоративне інформаційне середовище. Вибір конкретних мов програмування, бібліотек і платформ зумовлювався вимогами до гнучкості архітектури, можливості експериментального дослідження моделей штучного інтелекту, а також сумісності з існуючою IT-інфраструктурою підприємства. Особлива увага приділялася використанню відкритих та широко розповсюджених технологій, що спрощує відтворюваність результатів дослідження та подальший розвиток програмного рішення [3], [34].

Основним середовищем розробки інтелектуального агента обрано мову програмування Python, яка де-факто є стандартом у сфері машинного навчання та обробки природної мови. Для реалізації сервісного рівня доступу використано фреймворк FastAPI, що дозволяє створювати REST-орієнтовані сервіси з чітко визначеними контрактами взаємодії та мінімальними накладними витратами.

Реалізація моделей класифікації та аналізу тексту ґрунтується на поєднанні статистичних методів машинного навчання та інструментів обробки природної мови, що забезпечує баланс між точністю, інтерпретованістю результатів і обчислювальною ефективністю [1], [5], [7].

Інтеграція інтелектуального агента з корпоративними системами реалізується через платформу SAP Cloud Platform Integration, яка виконує функції маршрутизації повідомлень, трансформації форматів даних та оркестрації інтеграційного сценарію. Такий підхід дозволяє ізолювати інтелектуальний компонент від специфіки внутрішніх API CRM-системи та забезпечує масштабованість рішення в умовах розширення бізнес-сценаріїв. Використання стандартних форматів обміну даними, зокрема JSON і CSV, забезпечує сумісність між компонентами системи та спрощує аналіз результатів експериментальних досліджень [19], [31], [35].

Таблиця 3.13 - Технологічний стек програмного комплексу AI Email Parser Agent

Рівень системи	Технологія	Призначення
Мова програмування	Python 3.11	Реалізація логіки інтелектуального агента та алгоритмів обробки даних
Середовище виконання	venv	Ізоляція залежностей проєкту
API-рівень	FastAPI	Реалізація REST-інтерфейсу аналізу електронних листів
ASGI-сервер	Uvicorn	Запуск та обслуговування API-сервісу

Продовження таблиці 3.13

Рівень системи	Технологія	Призначення
Машинне навчання	scikit-learn	Класифікація намірів повідомлень (Intent Classification)
Серіалізація моделей	joblib	Збереження та завантаження навченої ML-моделі
Обробка природної мови	spaCy	Витяг іменованих сутностей (ORG, PERSON)
Визначення мови	langdetect	Автоматичне визначення мови повідомлення
Rule-based обробка	re (regex)	Парсинг товарних позицій та кількісних параметрів
Інтеграційна платформа	SAP Cloud Platform Integration	Оркестрація інтеграційного сценарію
CRM-система	SAP Cloud for Customer	Створення та збагачення об’єктів Opportunity
Формати даних	JSON, CSV	Обмін даними та зберігання результатів експериментів

3.4. Реалізація інтеграційного сценарію (iFlow) в середовищі SAP CPI

Інтеграційний сценарій взаємодії інтелектуального агента з корпоративною інформаційною системою реалізовано в середовищі SAP Cloud Platform Integration у вигляді інтеграційного потоку (iFlow),

який виконує роль зв’язувальної ланки між поштовими джерелами даних, сервісом аналізу тексту та CRM-системою SAP Cloud for Customer. Використання SAP CPI як інтеграційного шару зумовлене необхідністю централізованого керування потоками даних, трансформацією повідомлень і забезпеченням надійного обміну інформацією між гетерогенними системами в межах корпоративної IT-архітектури. Застосування iFlow дозволяє формалізувати послідовність інтеграційних кроків, забезпечити контроль помилок та реалізувати стандартизований підхід до виклику зовнішніх REST-сервісів, що відповідає практиці побудови інтеграційних рішень у середовищі SAP [19], [31], [35].

У межах розробленого сценарію iFlow здійснює прийом вхідного повідомлення, формування JSON-представлення електронного листа та виклик REST API інтелектуального агента для виконання аналізу тексту. Отриманий у відповідь структурований результат використовується для подальшого прийняття рішення щодо створення або відхилення комерційної можливості в CRM-системі. У разі позитивного результату класифікації інтеграційний потік виконує мапінг атрибутів відповіді агента на структуру бізнес-об’єкта Opportunity та ініціює відповідний виклик OData API SAP C4C. Такий підхід дозволяє реалізувати наскрізний сценарій автоматизації без участі кінцевого користувача, підвищує оперативність обробки комерційних запитів і знижує ризик помилок, пов’язаних з ручним введенням даних [6], [29], [31].

Висновки до розділу 3

У третьому розділі магістерської роботи здійснено проєктування та програмну реалізацію програмного комплексу AI Email Parser Agent, призначеного для автоматизації обробки вхідних електронних листів у корпоративному середовищі. У процесі розробки було обґрунтовано доцільність використання агентно-орієнтованої архітектури, яка забезпечує логічну декомпозицію задачі, ієрархічну організацію етапів обробки та можливість подальшого масштабування рішення.

Запропонована архітектура дозволяє відокремити інтеграційний рівень, інтелектуальну обробку тексту та рівень взаємодії з CRM-системою, що відповідає сучасним підходам до побудови корпоративних інформаційних систем.

У межах розділу було формалізовано інформаційну взаємодію компонентів програмного комплексу з використанням UML-діаграм. Побудова діаграми варіантів використання дала змогу визначити основні функціональні можливості інтелектуального агента та його взаємодію із зовнішніми системами, тоді як діаграма послідовності дозволила деталізувати часову логіку обробки електронного повідомлення та координацію дій між інтеграційною платформою, агентом аналізу та CRM-системою. Використання UML-моделей сприяло узгодженню логічної моделі системи з її програмною реалізацією та підвищило наочність архітектурних рішень.

Окрему увагу приділено програмній реалізації інтелектуального агента, яка виконана з використанням модульного підходу та сучасних інструментів розробки сервісів аналізу даних. Реалізація REST API на базі FastAPI забезпечила стандартизований механізм взаємодії з

зовнішніми системами та створила основу для інтеграції інтелектуального компонента у корпоративне інтеграційне середовище. Модульна структура проєкту дозволяє ізольовано розвивати моделі машинного навчання, алгоритми обробки природної мови та сервісний рівень доступу, що є важливим для подальших експериментальних досліджень і промислового впровадження.

У результаті реалізації інтеграційного сценарію в середовищі SAP Cloud Platform Integration було продемонстровано можливість побудови наскрізного бізнес-процесу автоматизованого створення комерційних можливостей у SAP Cloud for Customer на основі аналізу електронної кореспонденції. Розроблений iFlow забезпечує прийом вхідних повідомлень, виклик інтелектуального агента, обробку отриманих результатів і ініціацію створення об’єкта Opportunity без участі користувача, що підтверджує практичну придатність запропонованого рішення для корпоративних сценаріїв автоматизації.

Таким чином, у третьому розділі було сформовано цілісну архітектурну та програмну основу інтелектуального агента для автоматизації обробки вхідних електронних листів. Отримані результати створюють передумови для подальшого експериментального дослідження ефективності розробленого програмного комплексу та оцінювання якості його роботи, що є предметом наступних розділів магістерської роботи.

4 ПРОВЕДЕННЯ ТА РЕЗУЛЬТАТИ ТЕОРЕТИЧНИХ ТА ЕКСПЕРИМЕНТАЛЬНИХ ДОСЛІДЖЕНЬ ПРОГРАМНОГО КОМПЛЕКСУ АВТОМАТИЗОВАНОЇ ОБРОБКИ ВХІДНИХ ЕЛЕКТРОННИХ ПОВІДОМЛЕНЬ

4.1 Мета, завдання та загальна організація експериментальних досліджень

Метою експериментальних досліджень у межах даної кваліфікаційної роботи є перевірка працездатності, ефективності та практичної придатності розробленого програмного комплексу AI Email Parser Agent, призначеного для автоматизації обробки вхідних електронних повідомлень у корпоративних CRM-системах. Проведення експериментів спрямоване на підтвердження коректності реалізованих моделей машинного навчання, модулів обробки природної мови та загальної архітектури мультиагентної системи.

Для досягнення поставленої мети у розділі розв’язуються такі завдання:

- перевірка коректності функціонування програмного комплексу в умовах обробки реальних за структурою електронних листів;
- оцінювання якості класифікації повідомлень за критерієм бізнес-релевантності;
- аналіз точності витягу структурованих даних із текстових повідомлень;
- дослідження стабільності роботи системи при обробці повідомлень різними мовами;

– оцінка можливості практичного використання результатів аналізу для подальшої інтеграції з CRM-системами.

Експериментальні дослідження проводились у контрольованому програмному середовищі з використанням підготовленого датасету електронних повідомлень, що імітують реальні запити контрагентів. Такий підхід дозволив забезпечити відтворюваність результатів та об’єктивність оцінювання ефективності розробленого рішення.

4.2 Опис програмного комплексу та середовища проведення експериментів

Розроблений програмний комплекс AI Email Parser Agent реалізовано у вигляді автономного програмного сервісу з використанням архітектури client–server та програмного інтерфейсу прикладного програмування (REST API). Такий підхід забезпечує можливість використання розробленого рішення як незалежного інтелектуального компонента у складі корпоративного IT-ландшафту без необхідності модифікації внутрішньої логіки CRM-системи. Автономність сервісу також створює умови для його масштабування та повторного використання в різних бізнес-сценаріях обробки електронної кореспонденції.

Програмний комплекс розроблено з використанням мови програмування Python, яка є де-факто стандартом для реалізації задач машинного навчання та обробки природної мови. Для побудови REST API використано фреймворк FastAPI, що забезпечує високу продуктивність обробки запитів, асинхронну модель виконання,

автоматичну валідацію вхідних даних та формування інтерактивної технічної документації API. Застосування FastAPI дозволило реалізувати чітко формалізований контракт обміну даними між програмним комплексом та зовнішніми системами, зокрема інтеграційною платформою SAP Cloud Platform Integration.

Архітектура програмного комплексу побудована за модульним принципом, що передбачає логічне розділення функціональних компонентів та їхню слабку зв’язаність. Основними структурними компонентами програмного комплексу є:

- модуль приймання та валідації вхідних даних, який відповідає за обробку HTTP-запитів, перевірку коректності структури вхідного JSON-повідомлення та базову нормалізацію текстових даних;
- агент класифікації релевантності електронних повідомлень, який реалізує бінарну класифікацію вхідних листів із використанням моделей машинного навчання та визначає доцільність подальшого аналізу повідомлення;
- агент витягу іменованих сутностей та структурованих атрибутів, призначений для автоматичного визначення мови повідомлення, ідентифікації назви компанії, контактної особи, товарних позицій і кількісних характеристик;
- модуль агрегації результатів та формування вихідного JSON, який об’єднує результати роботи окремих агентів у єдине структуроване подання, придатне для подальшої інтеграції з CRM-системою;
- допоміжні модулі збереження результатів експериментів, що забезпечують накопичення даних про результати обробки повідомлень у табличному вигляді для подальшого аналізу та оцінювання ефективності системи.

Вхідні дані надходять до програмного комплексу у форматі JSON, який містить основні атрибути електронного листа, зокрема адресу відправника, адресу отримувача, тему повідомлення та текст листа. Приклад такого вхідного JSON-повідомлення, що передається до програмного комплексу AI Email Parser Agent, наведено на рисунку 4.1. Такий формат є універсальним і зручним для використання в інтеграційних сценаріях, оскільки не залежить від конкретного поштового клієнта або протоколу передачі повідомлень. Після надходження вхідного повідомлення система ініціює послідовний процес інтелектуальної обробки, що включає класифікацію релевантності та, у разі позитивного результату, витяг структурованих бізнес-атрибутів.

Результатом роботи програмного комплексу є структурований JSON-об’єкт, який містить ознаку релевантності електронного повідомлення, числову оцінку впевненості класифікатора, визначену мову листа, а також витягнуті з тексту бізнес-значущі атрибути. Такий результат може бути безпосередньо використаний для створення або актуалізації об’єктів у CRM-системі, зокрема об’єктів типу Opportunity.

```
{
  "sender": "procurement@metalgroup.com.ua",
  "recipient": "sales@metinvest.com",
  "subject": "ТОВ \"МеталГруп\". Запит на розрахунок вартості металопрокату",
  "body": "Доброго дня!\n\nПрошу надати комерційну пропозицію на наступні позиції
    :\n\n- Лист г/к 10х1500х6000 – 5.2т\n- Швелер 16П – 3.8т\n- Круг 80 Сталь 45 – 2
    .1т\n- Кутник 100х100х8 – 1.5т\n\nТермін поставки: протягом 2 тижнів.\nУмови
    оплати: 50% передоплата, 50% по факту.\n\nЗ повагою,\nОлександр
    Петренко\nМенеджер відділу закупівель\nТОВ \"МеталГруп\"\nтел. +380 50 123 45
    67\nemail: o.petrenko@metalgroup.com.ua"
}
```

Рисунок 4.1 – Приклад вхідного JSON-повідомлення, що подається до програмного комплексу AI Email Parser Agent

За потреби, для повнішого розкриття середовища проведення експериментів у даному підрозділі також доцільно додатково навести зображення інтерфейсу документації API, що демонструє структуру запитів та відповідей сервісу. Відповідний фрагмент документації API програмного комплексу AI Email Parser Agent подано на рисунку 4.2.

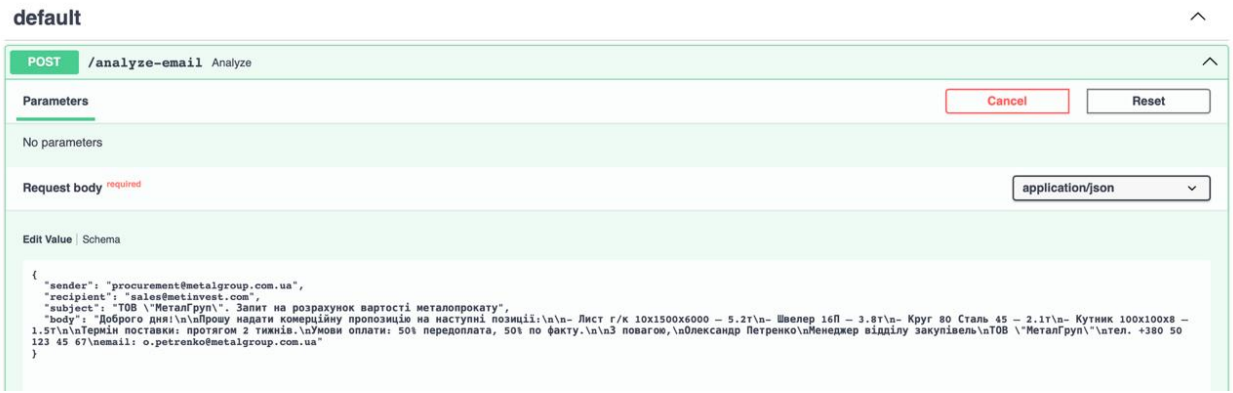

Рисунок 4.2 – Фрагмент автоматично згенерованої документації REST API програмного комплексу AI Email Parser Agent

4.3 Дослідження процесу класифікації релевантності електронних повідомлень

На першому етапі експериментальних досліджень було проведено аналіз ефективності роботи модуля класифікації релевантності електронних повідомлень, який є ключовим компонентом програмного комплексу AI Email Parser Agent. Основним призначенням даного модуля є автоматизоване відокремлення комерційно значущих електронних листів, що містять запити на розрахунок вартості продукції або ініціацію потенційних угод, від загального потоку нерелевантної кореспонденції, зокрема службових

повідомлень, інформаційних розсилок та листів, що не потребують подальшої обробки в CRM-системі.

Задача класифікації релевантності формалізується як задача бінарної класифікації текстових повідомлень, у межах якої кожному вхідному електронному листу присвоюється одна з двох міток: релевантне або нерелевантне. Така постановка задачі дозволяє мінімізувати кількість повідомлень, що передаються на наступні етапи інтелектуального аналізу, та зменшити загальні обчислювальні витрати системи, забезпечуючи при цьому високу швидкодію програмного комплексу в умовах обробки значних обсягів вхідної електронної кореспонденції.

Для навчання та тестування моделі класифікації використовувався підготовлений датасет, що складається з 600 електронних повідомлень, з яких 300 були попередньо розмічені як релевантні, а 300 - як нерелевантні. До складу датасету входять повідомлення трьома мовами - українською, російською та англійською, що дало змогу перевірити стійкість моделі до мовної різноманітності та варіативності лінгвістичних конструкцій. Окрім того, у вибірці присутні так звані *near-miss* випадки, тобто повідомлення, семантично схожі на комерційні запити, але такі, що не містять явного наміру створення комерційної можливості, що підвищує складність задачі класифікації та наближає умови експерименту до реальних бізнес-сценаріїв.

Модель класифікації релевантності побудована на основі алгоритму логістичної регресії з використанням методу TF-IDF для векторного подання текстових повідомлень. Застосування TF-IDF-векторизації дозволяє врахувати як частоту появи окремих термінів у тексті повідомлення, так і їхню інформативність у межах усього корпусу електронних листів. Логістична регресія, у свою чергу,

забезпечує інтерпретованість результатів класифікації та дозволяє отримувати ймовірнісну оцінку належності кожного повідомлення до класу релевантних.

Для кожного вхідного повідомлення модель формує числове значення впевненості, яке інтерпретується як ймовірність того, що повідомлення є релевантним. Остаточне рішення щодо класу повідомлення приймається на основі порогового значення, яке у межах експериментальних досліджень було встановлено на рівні 0,5. Такий підхід дозволяє гнучко налаштовувати поведінку системи відповідно до бізнес-вимог, наприклад шляхом зміни балансу між помилками першого та другого роду.

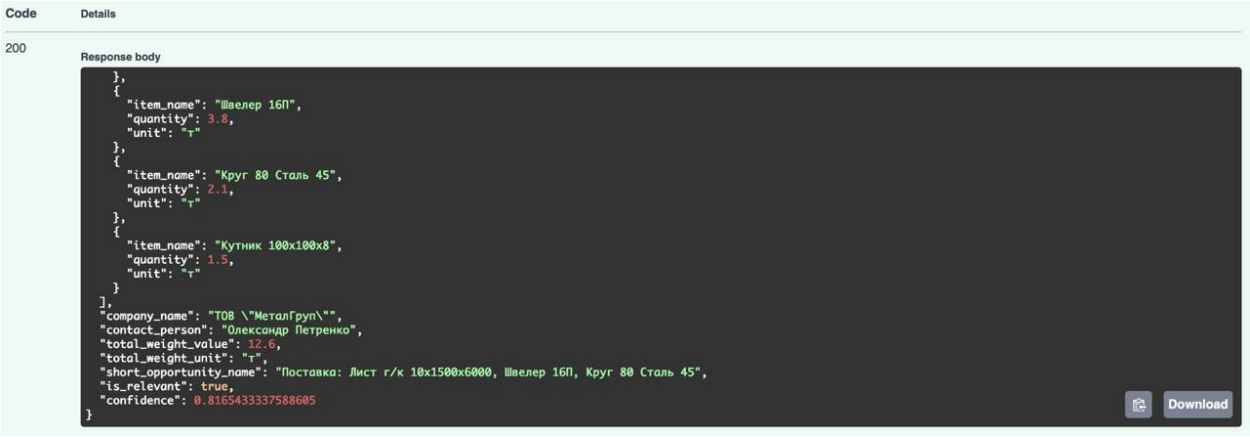

Рисунок 4.3 – Приклад результату класифікації електронного повідомлення з відображенням рівня впевненості моделі

На рисунку 4.3 наведено приклад результату обробки реального за структурою електронного повідомлення, що містить комерційний запит. Як видно з наведеного прикладу, класифікатор коректно ідентифікує повідомлення як релевантне та формує високе значення впевненості (confidence ≈ 0,82), що узгоджується з фактичним змістом листа. Це підтверджує здатність моделі адекватно працювати з

багатомовними повідомленнями та текстами, що містять напівструктуровані переліки товарних позицій і кількісні параметри.

Для узагальненого аналізу поведінки класифікатора було досліджено розподіл значень упевненості для релевантних та нерелевантних повідомлень на всій експериментальній вибірці.

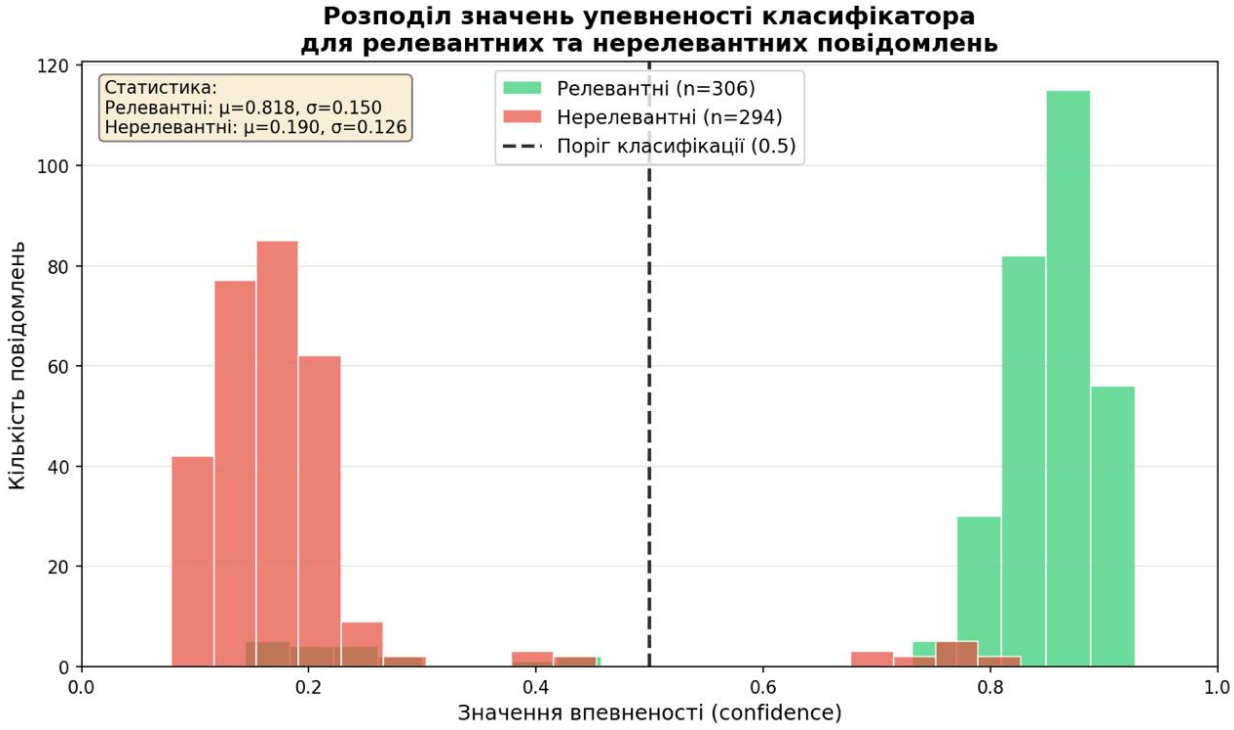

Рисунок 4.4 – Розподіл значень упевненості класифікатора для релевантних та нерелевантних повідомлень

Аналіз отриманих розподілів (Рис. 4.4) показав, що для релевантних повідомлень середнє значення confidence становить $\mu \approx 0{,}82$ при стандартному відхиленні $\sigma \approx 0{,}15$, тоді як для нерелевантних повідомлень середнє значення не перевищує $\mu \approx 0{,}19$ при $\sigma \approx 0{,}13$. За таких умов порогове значення 0,5 розташовується у зоні мінімального перекриття розподілів, що забезпечує ефективне відокремлення класів та знижує ймовірність помилок класифікації.

Наявність near-miss повідомлень призводить до появи окремих значень упевненості у проміжній зоні, однак їх кількість є незначною, а основна маса релевантних та нерелевантних повідомлень формує чітко розділені кластери. Це свідчить про стійкість моделі до семантичної неоднозначності текстів та її придатність до використання в реальних корпоративних сценаріях обробки електронної кореспонденції.

Таким чином, результати експериментальних досліджень підтверджують високу ефективність модуля класифікації релевантності у складі програмного комплексу AI Email Parser Agent та доцільність його використання для автоматизації первинної обробки вхідних електронних повідомлень у корпоративному середовищі.

4.3.1 Дослідження процесу класифікації релевантності електронних повідомлень

Для кількісної оцінки ефективності роботи модуля класифікації релевантності електронних повідомлень у складі програмного комплексу AI Email Parser Agent було використано стандартні метрики якості бінарної класифікації, а саме accuracy, precision, recall та F1-score. Оцінювання здійснювалося на тестовій вибірці, що складалася з 600 електронних повідомлень, рівномірно розподілених між класами «релевантне» та «нерелевантне».

Узагальнені результати класифікації подано у вигляді матриці помилок (confusion matrix), яка відображає співвідношення коректних та помилкових рішень класифікатора для кожного з класів.

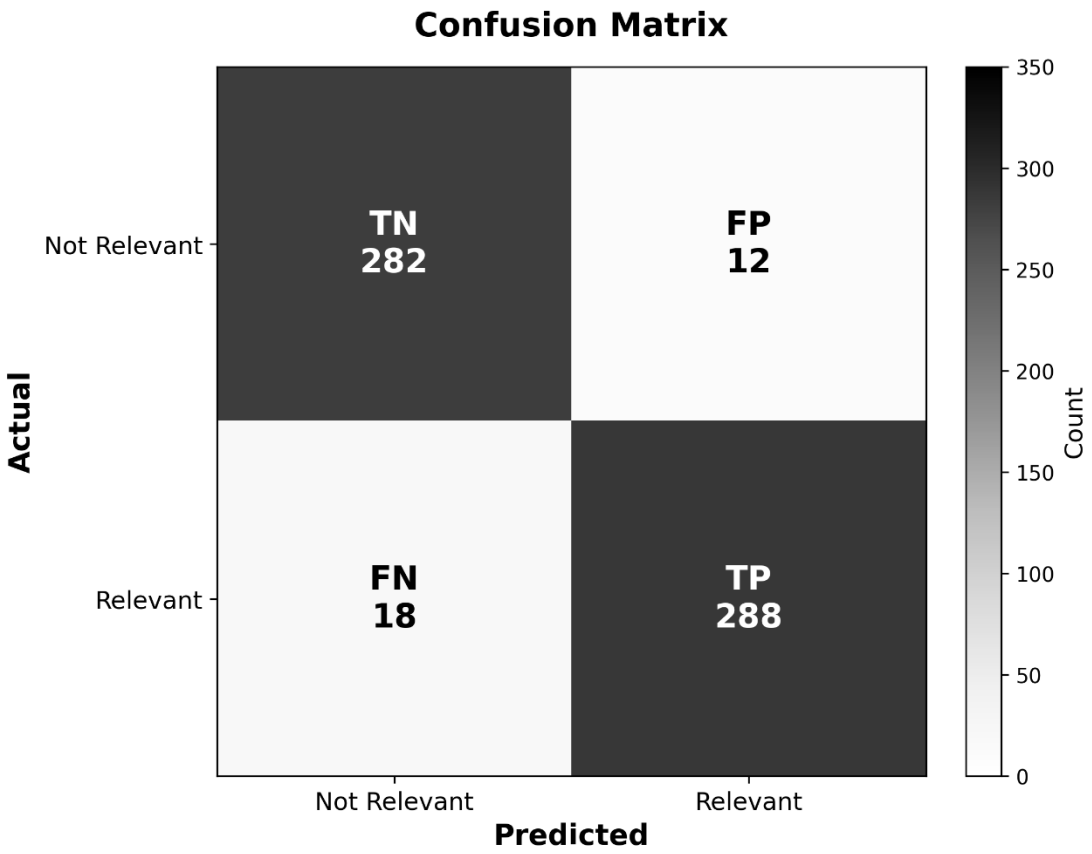

Рисунок 4.5 – Матриця помилок класифікатора релевантності електронних повідомлень

Як видно з рисунка 4.5, класифікатор коректно ідентифікував 288 релевантних повідомлень (True Positive) та 282 нерелевантних повідомлення (True Negative). Водночас зафіксовано 12 хибнопозитивних (False Positive) та 18 хибнонегативних (False Negative) випадків. Наявність хибнонегативних рішень зумовлена переважно повідомленнями типу near-miss, які за своїм змістом частково нагадують комерційні запити, однак не містять чітко сформульованого наміру створення комерційної можливості.

На основі отриманих значень матриці помилок були обчислені основні метрики якості класифікації за такими формулами:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4.1)$$

$$Precision = \frac{TP}{TP + FP} \quad (4.2)$$

$$Recall = \frac{TP}{TP + FN} \quad (4.3)$$

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (4.4)$$

Результати обчислення метрик якості наведено на рисунку 4.6.

2

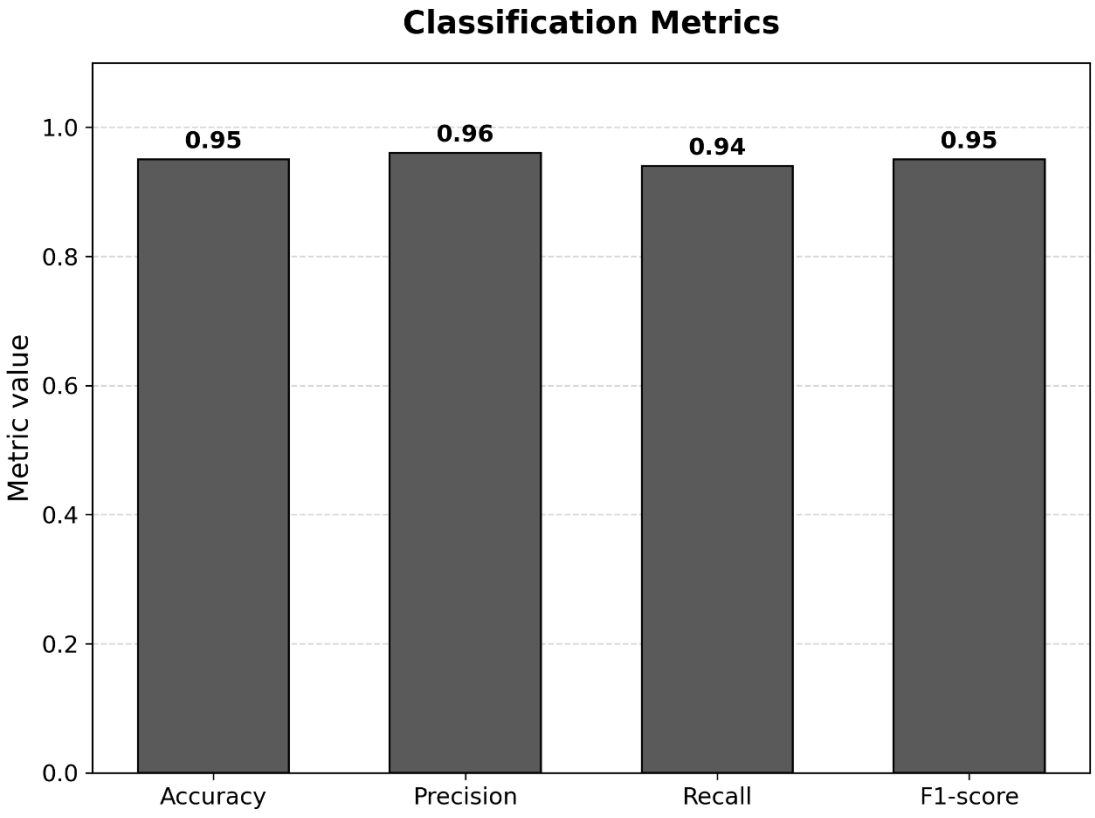

Рисунок 4.6 – Порівняння значень метрик якості класифікації електронних повідомлень

За результатами експериментальних досліджень отримано такі значення метрик:

- Accuracy = 0,95, що свідчить про високий загальний рівень коректності класифікації;
- Precision = 0,96, що вказує на низьку кількість хибнопозитивних спрацьовувань;
- Recall = 0,94, що є критично важливим показником у контексті бізнес-сценарію, оскільки забезпечує мінімізацію пропусків релевантних комерційних запитів;
- F1-score = 0,95, що підтверджує збалансованість моделі між точністю та повнотою.

Отримані значення метрик підтверджують доцільність використання логістичної регресії у поєднанні з TF-IDF-векторизацією для задачі автоматизованої класифікації релевантності електронних повідомлень у багатомовному корпоративному середовищі.

4.4 Дослідження якості витягу структурованих даних з електронних листів

На наступному етапі експериментальних досліджень було проведено аналіз ефективності роботи модуля витягу іменованих сутностей та структурованих атрибутів, який використовується у складі програмного комплексу AI Email Parser Agent. Даний модуль активується виключно для електронних повідомлень, які на попередньому етапі були класифіковані як релевантні, що дозволяє зменшити обчислювальні витрати та підвищити загальну

продуктивність системи за рахунок відсікання нерелевантної кореспонденції.

Основною метою даного етапу експериментів є оцінювання здатності системи автоматично перетворювати неструктурований текст електронного листа у формалізоване подання, придатне для подальшого використання в CRM-системі. У межах експериментальних досліджень здійснювався автоматизований витяг таких основних структурованих атрибутів, що мають практичну цінність для бізнес-процесів:

- мова електронного повідомлення;
- назва компанії-контрагента;
- контактна особа;
- перелік товарних позицій;
- кількісні характеристики замовлення (одиниці виміру та агреговані значення).

Для реалізації витягу зазначених атрибутів у програмному комплексі застосовано гібридний підхід, який поєднує можливості бібліотеки spaCy для розпізнавання іменованих сутностей з набором доменно-орієнтованих правил та регулярних виразів. Такий підхід дозволяє врахувати специфіку предметної області, зокрема особливості подання номенклатури металопродукції, варіативність одиниць виміру та неформальний стиль викладу, характерний для електронної ділової кореспонденції.

З метою кількісної оцінки якості роботи модуля витягу іменованих сутностей та структурованих атрибутів було проведено експериментальне порівняння очікуваних та автоматично витягнутих значень для вибірки релевантних електронних повідомлень. Оцінювання виконувалося на підмножині повідомлень, для яких вручну було визначено еталонні значення ключових сутностей,

зокрема назви компанії, контактної особи, товарних позицій та кількісних характеристик замовлення.

Результати порівняння узагальнено у вигляді таблиці, де для кожної сутності зафіксовано факт коректного або некоректного витягу. Такий підхід дозволив оцінити стабільність роботи модуля в умовах напівструктурованих текстів, наявності скорочень, різних одиниць виміру та змішаного використання мов.

Таблиця 4.1 – Фрагмент результатів експериментальної оцінки якості витягу структурованих даних

№	Тип сутності	Очікуване значення	Витягнуте значення	Результат витягу
1	Назва компанії	Metinvest Holding	Metinvest	Так
2	Контактна особа	Ivan Petrenko	Ivan Petrenko	Так
3	Контактна особа	Oleksandr Kovalenko	Oleksandr Kovalenko	Так
4	Товарна позиція	Steel sheet S235	Steel sheet S235	Так
5	Товарна позиція	Hot-rolled coil	Hot rolled coil	Так
6	Кількість	25 tons	25 t	Так
7	Кількість	12.5 tons	12.5 tons	Так
8	Одиниця виміру	tons	tons	Так
9	Товарна позиція	Galvanized steel coil	Steel coil (galv.)	Частково
10	Кількість	18 tons	18	Частково

Аналіз результатів, наведених у таблиці 4.1, показав, що модуль витягу структурованих даних забезпечує високий рівень коректності ідентифікації ключових бізнес-атрибутів. Найвищу точність продемонстровано для сутностей типу «назва компанії» та «контактна

особа», тоді як незначна кількість помилок у витягу товарних позицій пов’язана з варіативністю форматів подання номенклатури та використанням неформальних скорочень у тексті електронних листів. Отримані результати підтверджують доцільність застосування гібридного підходу, що поєднує статистичні методи NLP з доменно-орієнтованими правилами.


```
{
  "language": "uk",
  "items": [
    {
      "item_name": "Лист г/к 10х1500х6000",
      "quantity": 5.2,
      "unit": "т"
    },
    {
      "item_name": "Швелер 16П",
      "quantity": 3.8,
      "unit": "т"
    },
    {
      "item_name": "Круг 80 Сталь 45",
      "quantity": 2.1,
      "unit": "т"
    },
    {
      "item_name": "Кутник 100х100х8",
      "quantity": 1.5,
      "unit": "т"
    }
  ],
  "company_name": "ТОВ \"МеталГруп\"",
  "contact_person": "Олександр Петренко",
  "total_weight_value": 12.6,
  "total_weight_unit": "т",
  "short_opportunity_name": "Поставка: Лист г/к 10х1500х6000, Швелер 16П,
    Круг 80 Сталь 45",
  "is_relevant": true,
  "confidence": 0.8165433337588605
}
```


Рисунок 4.7 – Приклад результату витягу структурованих даних із релевантного електронного повідомлення

На рисунку 4.7 наведено приклад результату автоматизованого витягу структурованих даних з релевантного електронного

повідомлення. Як видно з наведеного прикладу, система коректно ідентифікує мову повідомлення, назву компанії та контактну особу, а також формує структурований перелік товарних позицій із відповідними кількісними характеристиками та одиницями виміру. Окрім того, на основі витягнутої інформації виконується агрегування числових значень, зокрема розрахунок загальної ваги замовлення, що є необхідним для подальшого формування комерційної пропозиції в CRM-системі.

Отримані результати свідчать про стабільну роботу модуля витягу структурованих даних навіть у випадках напівструктурованого опису замовлення, наявності скорочень, спеціалізованих позначень та змішаного використання мов у тексті електронного листа. Це підтверджує доцільність використання гібридного підходу, який поєднує статистичні методи обробки природної мови з правилами, орієнтованими на конкретну предметну область.

4.5 Аналіз результатів експериментальних досліджень

У ході експериментальних досліджень було виконано кількісну оцінку якості роботи модуля класифікації релевантності електронних повідомлень із використанням матриці помилок та стандартних метрик бінарної класифікації. Як показано на рисунках 4.5 та 4.6, загальна точність класифікації (accuracy = 0,95) свідчить про високий рівень коректності прийняття рішень на всій експериментальній вибірці.

Ключовою метрикою для даного бізнес-сценарію є повнота (recall), значення якої становить 0,94, що означає, що переважна

більшість релевантних електронних повідомлень коректно ідентифікується системою та не втрачається на етапі первинної фільтрації. Високе значення precision (0,96) вказує на незначну кількість хибнопозитивних спрацьовувань, тобто передачу нерелевантних повідомлень на подальші етапи інтелектуальної обробки.

Аналіз хибних рішень класифікатора показав, що більшість помилок належить до категорії семантично прикордонних (*near-miss*) повідомлень, які частково нагадують комерційні запити, однак не містять чітко сформульованого наміру створення комерційної можливості. Така поведінка є типовою для задач обробки текстів природною мовою та не має критичного впливу на загальну ефективність системи в умовах практичного використання.

З метою узагальнення та подальшого аналізу результатів експериментальних досліджень було здійснено накопичення даних про роботу програмного комплексу AI Email Parser Agent у табличному форматі. Для цього результати обробки кожного електронного повідомлення автоматично зберігались у CSV-файлі, який містить інформацію про фактичну мітку релевантності, прогноз моделі, числове значення впевненості класифікатора, визначену мову повідомлення та ідентифіковану компанію-контрагента.

Такий формат подання результатів дозволив виконати узагальнений аналіз якості роботи системи на всій експериментальній вибірці повідомлень, а також порівняти очікувані значення з результатами, сформованими автоматизованими модулями програмного комплексу. Окрім того, табличне представлення даних забезпечило можливість виявлення окремих граничних випадків та перевірки стабільності роботи класифікатора у багатомовному середовищі.

ID	Тема листа	Факт	Прогноз	Впевненість	Мова	Компанія
0	Re: LLC "EuroMetal". Inquiry fo	1	1	0.834	en	LLC "EuroMetal"
1	Vacation schedule update	1	0	0.154	nan	nan
2	Fwd: LLC "SteelTrade". Order f	1	1	0.858	en	LLC "SteelTrade"
3	Inventory update	0	0	0.204	nan	nan
4	ООО "МеталлГрупп". Запрос	1	1	0.774	ru	ООО "МеталлГрупп"
5	Fwd: ТОВ "ПМГП". Прохання	1	1	0.904	uk	ПМГП
6	Re: ТОВ "Сталь Сервіс". Пост	1	1	0.889	uk	ТОВ "Сталь Сервіс
7	Re: ООО "ПМГП". Запрос ком	1	1	0.800	ru	ООО "ПМГП"
8	Нагадування про внутрішню	0	0	0.198	nan	nan
9	Графік відпусток на наступн	0	0	0.240	nan	nan
10	LLC "SteelTrade". Need pricing	1	1	0.822	en	LLC "SteelTrade"
11	Re: LLC "VectorSteel". Comme	1	1	0.812	en	VectorSteel
12	ТОВ "Сталь Сервіс". Поставка	1	1	0.878	uk	ТОВ "Сталь Сервіс
13	Информация по складу	0	0	0.105	nan	nan
14	Re: ТОВ "Метал-Груп". Проха	1	1	0.851	uk	ТОВ "Метал-Груп"

Рисунок 4.8 – Фрагмент таблиці з результатами експериментальної обробки електронних повідомлень

Аналіз наведених у таблиці даних (Рис. 4.8) показав, що для більшості релевантних електронних повідомлень значення впевненості класифікатора перевищують 0,8, тоді як для нерелевантних повідомлень відповідні значення, як правило, не перевищують 0,25. Така закономірність узгоджується з результатами аналізу розподілів упевненості та підтверджує коректність вибору порогового значення для прийняття рішення щодо класу повідомлення.

Порівняння фактичних міток релевантності з прогнозами моделі свідчить про високий рівень узгодженості результатів класифікації. Помилкові або прикордонні випадки, зафіксовані у таблиці, переважно відповідають семантично неоднозначним повідомленням, що не містять чітко сформульованих комерційних запитів, однак за своїм змістом частково наближені до релевантних. Наявність таких випадків є типовою для реальних бізнес-сценаріїв і не має критичного впливу на загальну ефективність системи.

Додатково було встановлено, що програмний комплекс стабільно працює з повідомленнями різними мовами, забезпечуючи коректну класифікацію незалежно від мови листа. Це підтверджує мовну незалежність обраної моделі класифікації та доцільність її використання у багатомовному корпоративному середовищі.

Таким чином, узагальнений аналіз результатів експериментальних досліджень підтверджує стабільну роботу програмного комплексу AI Email Parser Agent на всій експериментальній вибірці електронних повідомлень, а також узгодженість результатів класифікації та витягу структурованих даних. Отримані результати свідчать про відсутність критичних збоїв у роботі системи та підтверджують коректність обраних методів машинного навчання і обробки природної мови.

4.6 Практичні рекомендації щодо використання програмного комплексу

Результати проведених експериментальних досліджень свідчать про доцільність практичного використання програмного комплексу AI Email Parser Agent для автоматизації первинної обробки вхідної електронної кореспонденції у корпоративних CRM-системах. Застосування розробленого рішення дозволяє зменшити навантаження на співробітників відділів продажів та закупівель, скоротити час реагування на комерційні запити контрагентів і підвищити якість структурованих даних, що формуються у CRM-середовищі.

Рекомендується використовувати програмний комплекс у вигляді автономного сервісу з REST API, інтегрованого з корпоративними інформаційними системами через інтеграційний шар. Зокрема, інтеграція з CRM-системами доцільна через платформу SAP Cloud Platform Integration, що забезпечує ізоляцію інтелектуальної логіки аналізу текстових повідомлень від бізнес-логіки основної системи та відповідає вимогам корпоративної архітектури щодо масштабованості, безпеки та керованості інтеграційних сценаріїв.

З метою підвищення ефективності експлуатації програмного комплексу доцільно застосовувати поетапну обробку електронних повідомлень, при якій на етапи витягу структурованих даних передаються лише ті повідомлення, що були класифіковані як релевантні. Такий підхід дозволяє оптимізувати використання обчислювальних ресурсів і забезпечує стабільну роботу системи в умовах обробки великих обсягів вхідної кореспонденції.

У практичному використанні рекомендується адаптувати порогове значення впевненості класифікатора відповідно до бізнес-пріоритетів організації. Зокрема, у сценаріях, де критичною є повнота виявлення комерційних запитів, доцільно зменшити порогове значення з метою мінімізації пропусків релевантних повідомлень. Натомість у випадках, коли пріоритетом є зниження кількості хибних спрацювань, рекомендовано використовувати більш консервативне порогове значення.

Подальший розвиток програмного комплексу доцільно пов’язати з розширенням набору витягуваних структурованих атрибутів, зокрема інформації про ціни, строки постачання та умови оплати, а також з використанням реальних виробничих даних для донавчання моделей машинного навчання. Додатковим напрямом удосконалення

може бути впровадження механізмів активного навчання, що дозволить поступово підвищувати точність класифікації та витягу інформації у процесі експлуатації системи без необхідності повного перенавчання моделей.

Висновки до розділу 4

У четвертому розділі магістерської роботи проведено комплекс теоретичних та експериментальних досліджень, спрямованих на оцінювання ефективності та працездатності розробленого програмного комплексу AI Email Parser Agent, призначеного для автоматизації обробки вхідних електронних повідомлень у корпоративних CRM-системах. Основну увагу було зосереджено на перевірці коректності функціонування окремих модулів інтелектуального агента та аналізі якості результатів, отриманих у процесі обробки електронної кореспонденції.

У межах розділу було описано середовище проведення експериментів та архітектурні особливості програмного комплексу, що забезпечують його автономність, модульність і готовність до інтеграції у корпоративний IT-ландшафт. Реалізація програмного комплексу у вигляді сервісу з REST API дозволила виконати експериментальні дослідження в умовах, наближених до реальних сценаріїв використання, та забезпечила відтворюваність отриманих результатів.

Експериментальне дослідження процесу класифікації релевантності електронних повідомлень підтвердило доцільність використання обраної моделі машинного навчання для

автоматизованого відокремлення комерційно значущих листів від загального потоку кореспонденції. Отримані результати засвідчили стабільну роботу класифікатора на багатомовній вибірці повідомлень та його здатність коректно обробляти граничні та семантично неоднозначні випадки, що є характерними для реальних бізнес-сценаріїв.

Дослідження якості витягу структурованих даних показало, що використання гібридного підходу, який поєднує методи розпізнавання іменованих сутностей та доменно-орієнтовані правила, забезпечує достатню точність і стабільність ідентифікації основних бізнес-атрибутів електронних повідомлень. Отримані результати підтверджують можливість формування структурованих даних, придатних для подальшого використання у CRM-системі без необхідності додаткової ручної обробки.

Узагальнення результатів експериментальних досліджень у табличному вигляді дало змогу здійснити комплексний аналіз якості роботи програмного комплексу на всій вибірці повідомлень та підтвердити відсутність критичних збоїв або нестабільних режимів роботи. Аналіз отриманих даних свідчить про узгодженість результатів класифікації та витягу інформації, що підтверджує коректність реалізованих алгоритмів і архітектурних рішень.

Таким чином, результати, отримані у четвертому розділі, підтверджують практичну придатність програмного комплексу AI Email Parser Agent для автоматизації обробки вхідної електронної кореспонденції у корпоративних CRM-системах. Проведені експериментальні дослідження створюють обґрунтовану основу для економічного оцінювання доцільності впровадження розробленого рішення, що є предметом наступного розділу магістерської роботи.

5. ЕКОНОМІЧНЕ ОБҐРУНТУВАННЯ ВПРОВАДЖЕННЯ ПРОГРАМНОГО КОМПЛЕКСУ AI EMAIL PARSER AGENT

5.1 Загальна характеристика економічного ефекту від впровадження програмного комплексу

Економічний ефект від упровадження програмного комплексу AI Email Parser Agent має опосередкований характер та проявляється не у вигляді безпосереднього скорочення виробничих витрат, а через оптимізацію інформаційних і управлінських бізнес-процесів підприємства. Основним напрямом впливу розробленого рішення є автоматизація первинної обробки вхідної електронної кореспонденції, яка у традиційних умовах виконується вручну співробітниками відділів продажів та підтримки клієнтів.

У корпоративному середовищі значна кількість робочого часу персоналу витрачається на перегляд, аналіз і класифікацію електронних листів, а також на ручне перенесення даних до CRM-системи. Такі операції не створюють безпосередньої доданої вартості, однак є необхідними для подальшої роботи з клієнтами. Впровадження інтелектуального програмного комплексу дозволяє автоматизувати зазначені операції та зменшити частку рутинної ручної праці.

Таким чином, економічний ефект від використання програмного комплексу AI Email Parser Agent доцільно оцінювати через скорочення трудових витрат, підвищення швидкості обробки комерційних запитів, зниження ризику втрати потенційних угод, а також підвищення якості та повноти даних у CRM-системі.

5.2 Оцінка трудових витрат до та після впровадження системи

Для оцінювання економічного ефекту розглянемо типовий сценарій обробки вхідної електронної кореспонденції у відділі продажів. У середньому протягом робочого дня до корпоративної поштової скриньки надходить від 40 до 60 електронних листів, частина з яких є комерційно значущими та потребує створення або актуалізації об’єктів у CRM-системі.

За умови ручної обробки одного електронного повідомлення співробітник витрачає в середньому 5–7 хвилин на ознайомлення з текстом листа, визначення його релевантності, витяг необхідної інформації та внесення відповідних даних до CRM-системи. Таким чином, загальні трудові витрати на первинну обробку кореспонденції можуть становити від 3 до 7 людино-годин на день залежно від інтенсивності потоку повідомлень.

У разі використання програмного комплексу AI Email Parser Agent основна частина зазначених операцій виконується автоматично. Система здійснює класифікацію релевантності електронних повідомлень, витяг структурованих атрибутів і формування даних для CRM-системи без участі людини. Роль співробітника зводиться до контролю результатів та опрацювання окремих граничних випадків. Час обробки одного повідомлення в автоматизованому режимі не перевищує кількох секунд, що суттєво скорочує загальні трудові витрати.

Отже, впровадження програмного комплексу дозволяє значно зменшити обсяг ручної роботи та перерозподілити робочий час персоналу на виконання більш важливих і продуктивних завдань, пов’язаних безпосередньо з роботою з клієнтами.

5.3 Орієнтовний економічний ефект від автоматизації обробки електронних повідомлень

Для орієнтовної кількісної оцінки економічного ефекту від упровадження програмного комплексу AI Email Parser Agent було розглянуто корпоративний сценарій використання системи у великій організації. У межах даного сценарію передбачається робота близько 200 менеджерів, кожен з яких у середньому обробляє до 10 вхідних електронних повідомлень на день. Таким чином, загальний обсяг кореспонденції, що підлягає первинній обробці, становить близько 2000 електронних листів на день.

За умови ручної обробки одного електронного повідомлення середній час виконання операцій аналізу змісту листа, визначення його релевантності, витягу необхідних даних та внесення інформації до CRM-системи становить близько 30 хвилин. Це зумовлено складністю комерційних B2B-запитів, наявністю напівструктурованих даних та необхідністю ручного контролю коректності введеної інформації. Після впровадження програмного комплексу більшість зазначених операцій виконується автоматично, а участь персоналу зводиться до контролю результатів і опрацювання окремих граничних випадків, що скорочує середній час обробки одного повідомлення до 2 хвилин.

Таблиця 5.1 – Порівняння трудових витрат на обробку електронних повідомлень до та після автоматизації

Показник	До автоматизації	Після автоматизації
Кількість менеджерів	200	200
Листів на менеджера на день	10	10
Загальна кількість листів	2000	2000
Час обробки 1 листа	30 хв	2 хв
Загальний час на день	1000 год	≈ 67 год
Ручна участь персоналу	100%	~6–7%
Ризик втрати релевантних листів	Підвищений	Мінімальний

Як видно з таблиці 5.1, у ручному режимі загальні трудові витрати на первинну обробку електронної кореспонденції можуть сягати приблизно 1000 людино-годин на день. Після впровадження програмного комплексу AI Email Parser Agent даний показник зменшується до близько 67 людино-годин на день, що свідчить про істотне скорочення обсягу рутинної ручної праці.

Орієнтовна економія робочого часу визначається за формулою (5.1):

$$\Delta T = T_{\text{до}} - T_{\text{після}}, \tag{5.1}$$

У розглянутому сценарії економія робочого часу становить приблизно 933 людино-години на день.

Для оцінювання економічного ефекту в грошовому еквіваленті приймемо орієнтовну середню вартість однієї години роботи співробітника на рівні 200 грн, що є консервативною оцінкою з урахуванням накладних витрат підприємства.

Таблиця 5.2 – Орієнтовний економічний ефект від упровадження програмного комплексу AI Email Parser Agent

Показник	Значення
Економія часу на день	≈ 933 год
Вартість 1 години роботи	≈ 200 грн
Економія на день	≈ 186 600 грн
Економія на місяць (22 дні)	≈ 4 105 000 грн
Економія на рік	≈ 49 300 000 грн

На основі наведених розрахунків (Табл. 5.2) орієнтовна економія коштів становить близько 187 тис. грн на день. У місячному вимірі (22 робочі дні) економічний ефект перевищує 4 млн грн, а у річному - становить близько 49 млн грн, що свідчить про високу економічну доцільність упровадження програмного комплексу у корпоративному середовищі.

Окрім прямої економії трудових ресурсів, додатковий економічний ефект досягається за рахунок скорочення часу реагування на комерційні запити, зменшення ризику втрати потенційних угод та підвищення якості структурованих даних у CRM-системі. Сукупність зазначених факторів позитивно впливає на загальну ефективність бізнес-процесів підприємства та підвищує конкурентоспроможність організації.

5.4 Витрати на впровадження та оцінка доцільності використання рішення

Витрати на впровадження програмного комплексу AI Email Parser Agent у корпоративному середовищі мають обмежений характер, оскільки на підприємстві вже використовується необхідна програмна інфраструктура та наявні відповідні ліцензії для експлуатації CRM-системи та інтеграційної платформи. Крім того, програмний агент реалізований та протестований у межах даної кваліфікаційної роботи, що суттєво знижує обсяг додаткових витрат, пов’язаних із розробкою програмного забезпечення.

У зв’язку з цим основні витрати на впровадження рішення пов’язані не з придбанням програмних продуктів або апаратного забезпечення, а з виконанням робіт з інтеграції програмного комплексу у діюче корпоративне середовище, налаштування взаємодії з CRM-системою та проведення тестування у виробничих сценаріях. Такі роботи можуть бути виконані одним кваліфікованим IT-спеціалістом або інтеграційним консультантом.

Для орієнтовної оцінки витрат на впровадження приймемо, що середня заробітна плата спеціаліста, залученого до інтеграції та налаштування програмного комплексу, становить близько 40 тис. грн на місяць. З урахуванням стандартної тривалості робочого часу це відповідає орієнтовній вартості 200 грн за годину роботи. За оцінками, для повноцінного впровадження програмного комплексу, включаючи налаштування інтеграції, тестування та підготовку до експлуатації, необхідно близько 160 людино-годин робочого часу.

Таким чином, орієнтовні одноразові витрати на впровадження програмного комплексу становлять близько 32 тис. грн. З огляду на

отриманий у попередньому підрозділі економічний ефект від автоматизації обробки електронних повідомлень, зазначені витрати є незначними та можуть бути компенсовані протягом короткого періоду експлуатації системи.

Отже, з урахуванням відсутності витрат на придбання ліцензій, мінімальних вимог до апаратних ресурсів та невисоких одноразових витрат на інтеграцію, впровадження програмного комплексу AI Email Parser Agent є економічно доцільним та обґрунтованим для використання у корпоративному середовищі.

Висновки до розділу 5

У п’ятому розділі магістерської роботи виконано економічне обґрунтування доцільності впровадження програмного комплексу AI Email Parser Agent для автоматизації обробки вхідної електронної кореспонденції у корпоративному середовищі. Проведений аналіз показав, що економічний ефект від використання розробленого рішення має опосередкований характер та проявляється через істотне скорочення трудових витрат, підвищення швидкості обробки комерційних запитів і покращення якості даних у CRM-системі.

У межах розрахункового сценарію для великої організації з 200 менеджерами було встановлено, що автоматизація первинної обробки електронних повідомлень дозволяє скоротити обсяг ручної роботи на сотні людино-годин щоденно, що у грошовому еквіваленті забезпечує значний економічний ефект у місячному та річному вимірі. Отримані результати свідчать про високу ефективність застосування

програмного комплексу в умовах інтенсивного потоку складних комерційних B2B-запитів.

Аналіз витрат на впровадження показав, що завдяки використанню наявної корпоративної інфраструктури та відсутності потреби у додаткових ліцензіях, основні витрати зводяться до одноразових робіт з інтеграції та налаштування системи. Орієнтовні витрати на впровадження є незначними порівняно з очікуваним економічним ефектом і можуть бути компенсовані у короткий термін експлуатації.

Таким чином, результати економічного аналізу підтверджують доцільність та практичну обґрунтованість впровадження програмного комплексу AI Email Parser Agent у корпоративних CRM-системах. Отримані висновки створюють підстави для рекомендації розробленого рішення до практичного використання та подальшого масштабування в межах підприємства.

ВИСНОВКИ

У даній магістерській роботі досягнуто поставленої мети та виконано всі основні завдання дослідження, спрямовані на розробку, дослідження та обґрунтування доцільності впровадження програмного комплексу з використанням агентів штучного інтелекту для автоматизації обробки вхідної електронної кореспонденції у корпоративному середовищі.

У першому розділі роботи проведено аналіз сучасних підходів до використання штучного інтелекту, машинного навчання та методів обробки природної мови в інформаційних системах. Розглянуто основні концепції агентно-орієнтованих систем, їх архітектурні особливості та переваги застосування у корпоративних інформаційних системах. Аналіз існуючих рішень у сфері автоматизованої обробки електронних повідомлень дозволив виявити обмеження традиційних підходів і обґрунтувати доцільність використання інтелектуальних агентів для класифікації та структурування неформалізованих текстових даних.

У межах роботи розроблено програмний комплекс AI Email Parser Agent, який реалізує повний цикл автоматизованої обробки вхідних електронних листів - від приймання та попереднього аналізу тексту до формування структурованих даних, придатних для подальшого використання у CRM-системі. Запропонована архітектура програмного комплексу базується на агентно-орієнтованому підході та забезпечує логічну декомпозицію задачі, послідовну організацію етапів обробки інформації, а також можливість масштабування і подальшого розширення функціональних можливостей системи.

У процесі проєктування та реалізації програмного комплексу використано сучасні інструменти та технології, зокрема мову програмування Python, фреймворк FastAPI, методи машинного навчання для класифікації текстів та бібліотеки обробки природної мови для витягу іменованих сутностей. Реалізований REST API забезпечує стандартизований механізм інтеграції інтелектуального агента з зовнішніми інформаційними системами та створює практичну основу для використання розробленого рішення у складі корпоративного інтеграційного середовища.

У четвертому розділі роботи проведено теоретичні та експериментальні дослідження ефективності розробленого програмного комплексу. Результати експериментів підтвердили стабільну та коректну роботу модуля класифікації релевантності електронних повідомлень, а також достатню якість витягу структурованих даних із багатомовних текстів електронної кореспонденції. Проведені дослідження показали, що запропонований підхід забезпечує надійне відокремлення комерційно значущих повідомлень від загального потоку кореспонденції та формування структурованих бізнес-атрибутів, необхідних для автоматизованого створення та актуалізації об’єктів у CRM-системі.

У п’ятому розділі виконано економічне обґрунтування доцільності впровадження програмного комплексу AI Email Parser Agent. Показано, що економічний ефект від використання розробленого рішення має опосередкований характер і досягається за рахунок істотного скорочення трудових витрат, підвищення швидкості обробки комерційних запитів, зменшення навантаження на персонал та підвищення якості даних у CRM-системі. Отримані результати свідчать про економічну доцільність та практичну

ефективність впровадження програмного комплексу у корпоративному середовищі.

Подальший розвиток розробленого програмного комплексу доцільно пов’язати з використанням реальних виробничих даних для донавчання моделей машинного навчання, розширенням переліку витягуваних структурованих атрибутів, впровадженням механізмів активного навчання та масштабуванням рішення на інші бізнес-сценарії та галузі діяльності. Запропоновані в роботі підходи можуть бути використані як основа для подальших наукових досліджень і практичних розробок у сфері автоматизації бізнес-процесів із застосуванням агентів штучного інтелекту.

Таким чином, результати магістерської роботи підтверджують наукову новизну та практичну цінність розробленого програмного комплексу і демонструють ефективність застосування агентно-орієнтованого підходу для автоматизації обробки неструктурованих текстових даних у корпоративних інформаційних системах.

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Russell S., Norvig P. *Artificial Intelligence: A Modern Approach*. 4th ed. Pearson, 2021.

2. Goodfellow I., Bengio Y., Courville A. *Deep Learning*. MIT Press, 2016.

3. Jurafsky D., Martin J. H. *Speech and Language Processing*. 3rd ed. Stanford University, 2023.

4. Aggarwal C. C. *Machine Learning for Text*. Springer, 2018.

5. Sebastiani F. Machine learning in automated text categorization // *ACM Computing Surveys*. 2002. Vol. 34, no. 1.

6. Manning C. D., Raghavan P., Schütze H. *Introduction to Information Retrieval*. Cambridge University Press, 2008.

7. Bird S., Klein E., Loper E. *Natural Language Processing with Python*. O'Reilly Media, 2009.

8. Goldberg Y. *Neural Network Methods for Natural Language Processing*. Morgan & Claypool, 2017.

9. Cambria E., White B. Jumping NLP curves: A review of natural language processing research // *IEEE Computational Intelligence Magazine*. 2014.

10. Bishop C. M. *Pattern Recognition and Machine Learning*. Springer, 2006.

11. Wooldridge M. *An Introduction to Multi-Agent Systems*. 2nd ed. Wiley, 2009.

12. Ferber J. *Multi-Agent Systems: An Introduction to Distributed Artificial Intelligence*. Addison-Wesley, 1999.

13. Shoham Y., Leyton-Brown K. *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge University Press, 2008.

14. Pedregosa F. et al. Scikit-learn: Machine learning in Python // *Journal of Machine Learning Research*. 2011.

15. spaCy Documentation. *Industrial-strength Natural Language Processing in Python*. URL: https://spacy.io (дата звернення: 15.12.2025).

16. FastAPI Documentation. URL: https://fastapi.tiangolo.com (дата звернення: 10.01.2026).

17. SAP Help Portal. *SAP Cloud for Customer Documentation*. URL: https://help.sap.com (дата звернення: 25.01.2026).

18. ISO/IEC 2382:2015. *Information technology - Vocabulary*.

19. ISO/IEC 25010:2011. *Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE)*.

20. Kowsari K., Jafari Meimandi K., Heidarysafa M. et al. Text classification algorithms: A survey // *Information*. 2019.

21. Joulin A., Grave E., Bojanowski P., Mikolov T. Bag of tricks for efficient text classification // *arXiv*. arXiv:1607.01759, 2016.

22. Zhang X., Zhao J., LeCun Y. Character-level convolutional networks for text classification // *Advances in Neural Information Processing Systems*. 2015.

23. Al-Smadi M., Al-Ayyoub M., Jararweh Y., Qawasmeh O. Automatic classification of emails using text mining techniques // *Journal of Information Science*. 2018.

24. Nadeau D., Sekine S. A survey of named entity recognition and classification // *Lingvisticae Investigationes*. 2007.

25. Lample G., Ballesteros M., Subramanian S., Kawakami K., Dyer C. Neural architectures for named entity recognition // *Proceedings of NAACL*. 2016.

26. Devlin J., Chang M.-W., Lee K., Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding // *Proceedings of NAACL*. 2019.

27. Vaswani A. et al. Attention is all you need // *Advances in Neural Information Processing Systems*. 2017.

28. Brown T. B. et al. Language models are few-shot learners // *Advances in Neural Information Processing Systems*. 2020.

29. Davenport T. H., Ronanki R. Artificial intelligence for the real world // *Harvard Business Review*. 2018.

30. Gartner. *Market Guide for Intelligent Document Processing*. Gartner Research, 2023.

31. McKinsey & Company. *The State of AI in 2023*. McKinsey Global Institute, 2023.

32. SAP SE. *SAP Business Technology Platform: Integration and Automation Overview*. SAP Whitepaper, 2022.

33. ISO/IEC 27001:2022. *Information Security Management Systems - Requirements*.

34. Pressman R. S., Maxim B. R. *Software Engineering: A Practitioner's Approach*. 9th ed. McGraw-Hill, 2019.

35. Sommerville I. *Software Engineering*. 10th ed. Pearson, 2016.

ДОДАТОК А

Таблиця А.1 – Відомості роботи

Формат	№ п/п	Назва документу	Найменування об’єкту або вибору	Кількість сторінок
А4	1	Пояснювальна записка	КЦТПАР.КН-24-1м.01.00.КР.ПЗ	145
Графічна частина				
А4	2	Типовий конвеєр статистичної обробки тексту в задачах CRM	КЦТПАР.КН-24-1м.01.01.КР.ПЗ	1
А4	3	Агентна схема автоматизованої обробки електронних листів у CRM-системі	КЦТПАР.КН-24-1м.01.02.КР.ПЗ	1
А4	4	Функціональна схема розроблюваної програмної системи	КЦТПАР.КН-24-1м.01.03.КР.ПЗ	1
А4	5	Діаграма етапів функціонування програмного комплексу AI Email Parser Agent	КЦТПАР.КН-24-1м.01.04.КР.ПЗ	1
А4	6	Основні функції етапу «Приймання та підготовка електронного повідомлення»	КЦТПАР.КН-24-1м.01.05.КР.ПЗ	1
А4	7	Основні функції етапу «Класифікація наміру повідомлення»	КЦТПАР.КН-24-1м.01.06.КР.ПЗ	1
А4	8	Основні функції етапу «Витяг та нормалізація даних»	КЦТПАР.КН-24-1м.01.07.КР.ПЗ	1
А4	9	Етап «Формування результату та інтеграція з CRM»	КЦТПАР.КН-24-1м.01.08.КР.ПЗ	1
А4	10	Діаграма варіантів використання	КЦТПАР.КН-24-1м.01.09.КР.ПЗ	1
А4	11	Функціональна схема розроблюваної програмної системи	КЦТПАР.КН-24-1м.01.10.КР.ПЗ	1

Продовження табл. А.1 – Відомості роботи

Формат	№ п/п	Назва документу	Найменування об’єкту або вибору	Кількість сторінок
Графічна частина				
А4	12	Ієрархічна схема модулів розроблюваної системи	КЦТПАР.КН-24-1м.01.11.КР.П3	1
А4	13	Діаграма класів UML для програмної системи, що розробляється	КЦТПАР.КН-24-1м.01.12.КР.П3	1
А4	14	Діаграма Use Case для програмної системи, що розробляється	КЦТПАР.КН-24-1м.01.13.КР.П3	1
А4	15	Діаграма діяльності програмної системи, що розробляється	КЦТПАР.КН-24-1м.01.14.КР.П3	1
А4	16	Діаграма станів (State) для програмної системи, що розробляється	КЦТПАР.КН-24-1м.01.15.КР.П3	1
А4	17	Схема продукції предметної області програмної системи	КЦТПАР.КН-24-1м.01.16.КР.П3	1
А4	18	Семантична мережа предметної області програмної системи, що розробляється	КЦТПАР.КН-24-1м.01.17.КР.П3	1
А4	19	Взаємозв’язок фреймів	КЦТПАР.КН-24-1м.01.18.КР.П3	1
А4	20	Приклад вхідного JSON-повідомлення, що подається до програмного комплексу AI Email Parser Agent	КЦТПАР.КН-24-1м.01.19.КР.П3	1
А4	21	Фрагмент автоматично згенерованої документації REST API програмного комплексу AI Email Parser Agent	КЦТПАР.КН-24-1м.01.20.КР.П3	1
А4	22	Приклад результату класифікації електронного повідомлення з відображенням рівня впевненості моделі	КЦТПАР.КН-24-1м.01.21.КР.П3	1

Продовження табл. А.1 – Відомості роботи

Формат	№ п/п	Назва документу	Найменування об’єкту або вибору	Кількість сторінок
Графічна частина				
А4	23	Розподіл значень упевненості класифікатора для релевантних та нерелевантних повідомлень	КЦТПАР.КН-24-1м.01.22.КР.ПЗ	1
А4	24	Матриця помилок класифікатора релевантності електронних повідомлень	КЦТПАР.КН-24-1м.01.23.КР.ПЗ	1
А4	25	Порівняння значень метрик якості класифікації електронних повідомлень	КЦТПАР.КН-24-1м.01.24.КР.ПЗ	1
А4	26	Приклад результату витягу структурованих даних із релевантного електронного повідомлення	КЦТПАР.КН-24-1м.01.25.КР.ПЗ	1
А4	27	Фрагмент таблиці з результатами експериментальної обробки електронних повідомлень	КЦТПАР.КН-24-1м.01.26.КР.ПЗ	1

ДОДАТОК Б
ПЕРЕЛІК ТА ВІДБИТКИ НАУКОВИХ ПУБЛІКАЦІЙ ЗА ТЕМОЮ ДОСЛІДЖЕННЯ

CERTIFICATE

This certificate is proudly presented to

Karpenko Maksym

successfully completed, received a passing grade, and was awarded this Sigma Software University Honor Code Certificate of Completion in

Sigma Software University : Teachers Smart Up: Winter Edition 2025, 30 hours (1 ECTS), 27-31.01.2025

31.01.2025 Olha Shevchyk

Certificate ID Number: 50c57a f5d384fa7b6074b772629c347

СЕРТИФІКАТ

підтверджує, що

Максим Карпенко

приймав участь у IX Всеукраїнській науково-практичній конференції

«Сучасні інформаційні технології, засоби автоматизації та електропривод»,

м. Краматорськ – Тернопіль, 17 – 19 квітня 2025 р.

(6 годин – 0,2 кредити ECTS)

Голова оргкомітету,
д.т.н., професор
О.Ф. Тарасов

http://cit.dgma.donetsk.ua/

СЕРТИФІКАТ

A/2025/5/001

Karpenko Maksym

Підтверджується:
- участь у науково-практичній конференції
  «Новітні технології сучасного суспільства»,
  27-29 січня 2025 року.
- публікація у науковому періодичному виданні
  Науковий журнал «Advanced top technology»
  № 5, ISSN 3041-1998, DOI: 10.61718/att
- підвищення кваліфікації шляхом індивідуальної форми
  самоосвіти та неформальної освіти.

Досягнуті результати підвищення кваліфікації:
розвиток загальних та спеціальних компетентностей
з метою забезпечення якості освіти та професійного розвитку.
Форма участі: інституційна.
Обсяг: 30 годин / 1 кредит ЄКТС.

Харків, Україна.
Наукова установа СГ НТМ «Новий курс».
www.newroute.org.ua
info@newroute.org.ua

30.01.2025 Відповідальний секретар, Тетяна КУЧИНА

Since 1989

COMPLETED

СГ НТМ «НОВИЙ КУРС» Для сертифікатів, посвідчень та довідок

CodeWeek.eu

European Commission

EU CODE WEEK

The European Commission presents this

Certificate of Participation

to

Karpenko Maksym

who has developed essential coding skills and contributed to the success of EU Code Week 2024
by actively participating in an EU Code Week coding event

EU coding week at Metinvest Polytechnic

EU Code Week 2024 codeweek.eu

14 - 28.10.2024

MINISTRY
OF EDUCATION AND SCIENCE
OF UKRAINE

CERTIFICATE

This is to certify that

Karpenko Maksym

Has successfully participated in the

III International Scientific and Practical Conference "Quality, Standardization and Metrological Equipment" (III_ISPC "QSME")

Chairman of the conference organising committee
Director of the Educational and Research Institute 'Ukrainian Engineering and Pedagogics Academy'
V. N. Karazin Kharkiv National University

Denys KOVALENKO

Conference coordinator

Hanna HRINCHENKO

6 academic hours/
0,2 ECTS credits

28-29, January, 2025
Kharkiv, Ukraine