Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет інформаційно-аналітичних технологій та менеджменту
(повна назва)

Кафедра прикладної математики
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА
Пояснювальна записка

рівень вищої освіти другий (магістерський)

Побудова GenAI чат-боту
в захищеному середовищі для роботи з
конфіденційними даними
(тема)

Виконав:
здобувач _2_ року навчання, групи _САУм-23-2
Харченко І.В.
(прізвище, ініціали)
Спеціальність
124 Системний аналіз
(код і повна назва спеціальності)
Тип програми _освітньо-професійна
(освітньо-професійна або освітньо-наукова)
Освітня програма
Системний аналіз і управління
(повна назва освітньої програми)
Керівник асист. Луханін В.С.
(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри ПМ _____ Сидоров М.В.
(підпис) (прізвище, ініціали)

2025 р.

Харківський національний університет радіоелектроніки

Факультет _інформаційно-аналітичних технологій та менеджменту
Кафедра _прикладної математики
Рівень вищої освіти _другий (магістерський)
Спеціальність _124 Системний аналіз
(код і повна назва)
Тип програми _освітньо-професійна
(освітньо-професійна або освітньо-наукова)
Освітня програма _Системний аналіз і управління
(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри ПМ _____
(підпис)
“ _25_ ” _листопада_ 2024 р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

здобувачеві _Харченку Івану Вікторовичу
(прізвище, ім’я, по батькові)

1. Тема роботи _Побудова GenAI чат-боту в захищеному середовищі для роботи з конфіденційними даними

затверджена наказом по університету від 22 листопада 2024 р. № 1228 Ст

2. Термін подання здобувачем роботи до екзаменаційної комісії 6 січня 2025 р.

3. Вихідні дані до роботи _модель хмарної мережевої інфраструктури з використанням програмного забезпечення Terraform

4. Перелік питань, що потрібно опрацювати в роботі
1. Системний аналіз предметної області
2. Вибір і обґрунтування методу розв’язання
3. Програмна реалізація
4. Результати обчислювального експерименту
5. Аналіз можливих застосувань

5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій

1. Актуальність теми роботи
2. Постановка задачі
3. Системний аналіз предметної області
4. Метод чисельного аналізу
5. Результати обчислювального експерименту

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
1	Підбір та вивчення технічної літератури за темою роботи	25 листопада – 1 грудня 2024 р.	виконано
2	Вибір та обґрунтування методу	2 – 8 грудня 2024 р.	виконано
3	Розробка алгоритму і програми	9 – 22 грудня 2023 р.	виконано
4	Проведення аналітичних досліджень та розрахунків	23 – 29 грудня 2024 р.	виконано
5	Робота над текстом пояснювальної записки	30 грудня 2024 р. – 9 січня 2025 р.	виконано
6	Представлення роботи на рецензію в ЕК	10 січня 2025 р.	виконано

Дата видачі завдання _25_ листопада 2024 р.

Здобувач _____
(підпис)

Керівник роботи _____ асист. Луханін В.С.
(підпис) (посада, прізвище, ініціали)

РЕФЕРАТ

Пояснювальна записка: 68 с., 6 табл., 8 рис., 1 дод., 10 джерел.

AWS BEDROCK, CLAUDE, КОНФІДЕНЦІЙНІ ДАНІ, ІНФОРМАЦІЙНА БЕЗПЕКА, TERRAFORM, ГЕНЕРАТИВНІ МОВНІ МОДЕЛІ.

Об’єкт дослідження – процес обробки конфіденційних даних за допомогою генеративних мовних моделей у захищеному середовищі.

Мета роботи – розробка GenAI чат-бота, який забезпечує безпечну обробку конфіденційних запитів, дотримуючись високих стандартів інформаційної безпеки.

Методи дослідження – системний аналіз, архітектурне моделювання, методи шифрування, управління хмарною інфраструктурою за допомогою Terraform.

У кваліфікаційній роботі розглянуто проблему використання генеративного AI (GenAI) для роботи з конфіденційними даними. Проведено системний аналіз задачі, обрано оптимальну мовну модель Claude у складі AWS Bedrock, та розроблено захищену інфраструктуру з використанням Terraform. Запропонована система забезпечує високий рівень конфіденційності, продуктивності та релевантності обробки запитів.

Створено програмну архітектуру, яка відповідає міжнародним стандартам, включаючи GDPR, HIPAA та CCPA. Проведено тестування системи у середовищі AWS. Результати демонструють точність відповідей на рівні 92% та релевантність у 95%, із середнім часом обробки текстових запитів 1,8 с.

Система має потенціал використання у сферах охорони здоров’я, фінансів та державного управління, де безпека інформації є критичною. Запропоновані підходи дозволяють адаптувати систему до галузевих специфікацій, інтегрувати її з корпоративними інструментами (CRM, ERP) та масштабувати для великих обсягів даних.

ABSTRACT

Introductory note: 68 pages, 6 tables, 8 figures, 1 appendix, 10 sources.

AWS BEDROCK, CLAUDE, CONFIDENTIAL DATA, INFORMATION SECURITY, TERRAFORM, GENERATIVE LANGUAGE MODELS.

Object of research – the process of handling confidential data using generative language models in a secure environment.

Purpose of work – to develop a GenAI chatbot that ensures secure processing of confidential queries while adhering to high information security standards.

Methods of research – system analysis, architectural modeling, encryption techniques, and cloud infrastructure management using Terraform.

The qualification work addresses the issue of using generative AI (GenAI) for handling confidential data. A systematic analysis of the task was conducted, the Claude language model within AWS Bedrock was chosen as the optimal solution, and a secure infrastructure was developed using Terraform. The proposed system ensures a high level of confidentiality, performance, and query relevance.

A software architecture was created that complies with international standards, including GDPR, HIPAA, and CCPA. The system was tested in the AWS environment. The results demonstrate a response accuracy of 92% and relevance of 95%, with an average processing time of 1.8 seconds for text queries.

The system has potential applications in healthcare, finance, and public administration, where information security is critical. The proposed approaches enable the system to be adapted to industry-specific requirements, integrated with corporate tools (CRM, ERP), and scaled for large data volumes.

ЗМІСТ

ПЕРЕЛІК СКОРОЧЕНЬ, УМОВНИХ ПОЗНАК, ОДИНИЦЬ І ТЕРМІНІВ

AWS – Amazon Web Services;

AWS Bedrock – інфраструктура для генеративного AI на платформі Amazon Web Services;

AWS IAM – служба управління ідентифікацією та доступом в AWS;

AWS PrivateLink – технологія для приватного доступу до хмарних сервісів AWS;

AWS S3 – Simple Storage Service, хмарне сховище даних від AWS;

S3-бакет – хмарне сховище даних у AWS S3;

AWS CloudWatch – система моніторингу та управління AWS;

AWS KMS – служба управління ключами в AWS;

AWS Lambda – обчислювальна платформа без серверів в AWS;

AWS Access Analyzer – інструмент для аналізу доступу до ресурсів AWS;

AWS CloudTrail – служба журналювання подій в AWS;

ARN – Amazon Resource Name, унікальний ідентифікатор ресурсу AWS;

EC2 – Elastic Compute Cloud, хмарні віртуальні сервери від AWS;

RDS – Relational Database Service, реляційна база даних від AWS;

VPC Endpoint Policies – політики кінцевих точок віртуальних приватних хмар;

VPC Flow Logs – журнали мережевого трафіку в AWS VPC;

Route Tables – таблиці маршрутизації в мережах VPC;

SSE – серверне шифрування (Server-Side Encryption);

VPC – Virtual Private Cloud, віртуальна приватна мережа;

GDPR – General Data Protection Regulation, Загальний регламент із захисту даних (ЄС);

HIPAA – Health Insurance Portability and Accountability Act, Закон про переносимість та відповідальність медичного страхування (США);

CCPA – California Consumer Privacy Act, Закон Каліфорнії про конфіденційність споживачів;

LLM – Large Language Model, велика мовна модель;

GPT – Generative Pre-trained Transformer, генеративна попередньо навчена трансформерна модель;

BERT – Bidirectional Encoder Representations from Transformers, двонаправлене кодування із трансформерами;

Terraform – інструмент для управління інфраструктурою як кодом;

HCL – HashiCorp Configuration Language, декларативна мова конфігурації Terraform;

IaC – Infrastructure as Code, інфраструктура як код;

Anthropic – компанія-розробник генеративного AI;

Claude – генеративна мовна модель від Anthropic;

Microsoft Azure – хмарна платформа від Microsoft;

Google Cloud – хмарна платформа від Google;

Google Gemini – генеративна мовна модель Google;

PaLM – Pathways Language Model, мовна модель від Google;

OpenAI ChatGPT – чат-бот на основі великих мовних моделей від OpenAI;

GPT-3.5, GPT-4 та GPT-4 Turbo – покоління генеративних моделей від OpenAI;

IDEF0 – метод моделювання функцій систем;

DFD – Data Flow Diagram, діаграма потоків даних;

AES-256 – стандарт шифрування Advanced Encryption Standard з довжиною ключа 256 біт;

TLS 1.3 – Transport Layer Security версії 1.3, протокол шифрування для передачі даних;

SHA-256 – алгоритм хешування Secure Hash Algorithm 256-біт;

Zero Trust – концепція інформаційної безпеки «нульової довіри»;

Фаєрвол – програмне або апаратне забезпечення для захисту мережі;

IDS/IPS – Intrusion Detection and Prevention Systems, системи виявлення та запобігання вторгнень;

MFA – Multi-Factor Authentication, багатофакторна автентифікація;

HSM – Hardware Security Module, апаратний модуль безпеки;

CIDR-блок – блок адрес за схемою Classless Inter-Domain Routing;

ACL – Access Control List, список контролю доступу;

NAT – Network Address Translation, трансляція мережевих адрес;

URL-адреса – уніфікований локатор ресурсів;

NAT Gateway – шлюз для трансляції мережевих адрес.

ВСТУП

Актуальність теми. Розповсюдження генеративного штучного інтелекту значно трансформувало підходи до автоматизації бізнес-процесів, взаємодії з користувачами та обробки великих обсягів даних. Зокрема, чат-боти, засновані на великих мовних моделях (LLM), забезпечують ефективні способи автоматизації комунікацій у різних галузях, таких як медицина, банківська справа, освіта та державне управління. Однак, впровадження таких систем у середовищах з конфіденційними даними викликає низку викликів, пов’язаних із забезпеченням інформаційної безпеки та запобіганням витокам даних.

Забезпечення конфіденційності даних та захищеного середовища для роботи з чутливою інформацією є пріоритетом для багатьох міжнародних компаній і дослідницьких організацій. Сучасні технології, такі як AWS Bedrock та інструменти управління інфраструктурою, наприклад, Terraform, відкривають можливості для створення захищених середовищ, що відповідають високим стандартам безпеки. Світові лідери у сфері хмарних обчислень впроваджують стратегії інтеграції генеративного штучного інтелекту з безпечними платформами для обробки конфіденційних даних.

Необхідність захисту даних у секторах, що працюють із конфіденційною інформацією, таких як банківська справа, охорона здоров’я та юридичні послуги, зумовлює важливість розробки GenAI чат-ботів, які поєднують функціональність сучасних мовних моделей із безпекою середовища, в якому вони працюють.

Мета і завдання кваліфікаційної роботи. Метою кваліфікаційної роботи є створення генеративного AI (GenAI) чат-боту, який працюватиме в захищеному середовищі для обробки конфіденційної інформації (без ризиків/з мінімізацією ризиків витоку даних). Для досягнення поставленої мети необхідно виконати наступні завдання:

– провести огляд і аналіз сучасного стану задачі «використання GenAI чат-ботів для роботи з конфіденційними даними»;

– розробити архітектуру захищеного середовища з використанням сервісу AWS Bedrock;

– реалізувати GenAI чат-бот з урахуванням вимог конфіденційності;

– оцінити ефективність запропонованого рішення.

Об'єктом дослідження є процес взаємодії з конфіденційними даними при використанні генеративного AI.

Предметом дослідження є використання GenAI чат-ботів з конфіденційною інформацією та забезпечення їхньої роботи в захищеному середовищі.

Методи дослідження. У кваліфікаційній роботі використовуються методи системного аналізу, проєктування захищених інфраструктур, моделі генеративного штучного інтелекту та сучасні засоби управління інфраструктурою, такі як Terraform.

Публікації. Результати, отримані у кваліфікаційній роботі, було представлено на 13-ій Міжнародній науково-технічній конференції «Інформаційні системи та технології ICT-2024» (м. Харків, 26-27 листопада 2024 р.) [1].

1 СИСТЕМНИЙ АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ТА ПОСТАНОВКА ЗАДАЧ ДОСЛІДЖЕННЯ

1.1 Системний аналіз задачі використання GenAI чат-ботів з конфіденційною інформацією

1.1.1 Вербальна модель системи

Об’єктом аналізу є інформаційна система з GenAI чат-ботом, призначена для роботи з конфіденційними даними в захищеному середовищі. Основна задача системи полягає у забезпеченні автоматизації інтерактивної взаємодії користувачів із системою, що використовує LLM для генерації релевантних відповідей на запити користувачів, при цьому суворо дотримуючись стандартів інформаційної безпеки та конфіденційності.

Призначення системи: забезпечення безпечного та конфіденційного діалогу користувачів із системою для вирішення професійних або організаційних задач, включаючи роботу з чутливими даними (персональною інформацією, фінансовими звітами, медичними даними тощо) за допомогою інтеграції генеративних мовних моделей та сучасних технологій шифрування й управління доступом.

Мета системи: створення безпечного середовища для обробки, аналізу та генерації інформації з використанням генеративних мовних моделей, інтегрованих у хмарну інфраструктуру, яка забезпечує відповідність регуляторним вимогам, таким як GDPR, HIPAA, CCPA, а також автоматизацію та масштабованість за допомогою таких технологій, як AWS Bedrock і Terraform.

Проведемо класифікацію системи.

За походженням система є штучною, оскільки створена людиною для обробки конфіденційних даних із застосуванням генеративних мовних моделей.

За об’єктивністю існування система є абстрактною, адже представлена як концептуальна модель та програмне забезпечення, інтегроване в хмарну інфра-

структуру.

За природою система належить до соціотехнічних, оскільки об’єднує технічні компоненти, такі як мовні моделі та хмарні сервіси, із людським фактором – користувачами та адміністраторами.

За централізованістю система є централізованою, адже її управління здійснюється через хмарну інфраструктуру, яка виступає центром обробки запитів.

За розмірністю система є багатовимірною, оскільки має численні входи, обробляє їх паралельно та видає відповідні результати.

За однорідністю структурних елементів система є гетерогенною, оскільки складається з різних компонентів, таких як мовні моделі, системи безпеки та інфраструктура шифрування, кожен з яких виконує унікальну функцію.

За лінійністю система є нелінійною, оскільки її поведінка залежить від складних взаємодій між компонентами, зокрема мовними моделями, системами безпеки та регуляторними обмеженнями.

За цільовою орієнтацією система є цілеспрямованою, оскільки її мета визначається внутрішніми функціями для забезпечення безпеки даних і конфіденційного діалогу.

За складністю система є складною, оскільки характеризується взаємозалежністю компонентів і складними алгоритмами взаємодії, такими як генеративні мовні моделі та механізми шифрування.

За ступенем детермінованості система є недетермінованою (стохастичною), оскільки використовує стохастичні алгоритми, притаманні генеративним мовним моделям.

За взаємодією із зовнішнім середовищем система є відкритою, адже активно обмінюється даними з користувачами та інфраструктурою.

За способом організації система є ієрархічною, оскільки структурована як багаторівнева модель, де мовні моделі, інфраструктура та безпекові механізми підпорядковуються централізованому управлінню.

За способом управління система має комбіноване управління, оскільки частина функцій, таких як обробка даних, автоматизована, а частина, наприклад

регуляторні налаштування, вимагає втручання людини.

За статичністю система є динамічною, адже її стан змінюється в реальному часі залежно від запитів користувачів і зовнішніх умов.

1.1.2 Морфологічний опис системи

Морфологічний опис системи «GenAI чат-бот у захищеному середовищі» відображає її структуру, складові частини, взаємодію із зовнішнім середовищем та основні компоненти, необхідні для забезпечення функціонування. Система представляється у вигляді моделі типу «чорна скринька», де вхідні дані – це текстові запити користувача, які можуть містити чутливу або конфіденційну інформацію, таку як персональні дані, фінансові звіти або медична інформація. На виході система генерує релевантну відповідь, створену за допомогою генеративної мовної моделі (LLM), з урахуванням контексту та точності запиту. Внутрішній склад системи користувачем не досліджується, а увага зосереджується на її межах, які забезпечують цілісність і конфіденційність. Ці межі включають інтеграцію з хмарною інфраструктурою (AWS Bedrock), яка забезпечує безпечне середовище виконання; політики доступу та управління правами користувачів (IAM), що обмежують доступ до чутливих даних; а також використання внутрішнього шифрування та захищених точок взаємодії (AWS PrivateLink), які гарантують захист інформації на всіх етапах обробки.

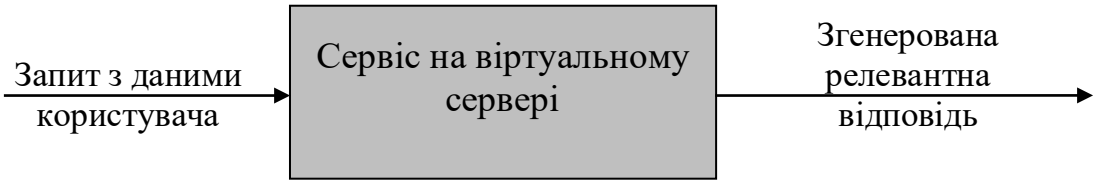

Рисунок 1.1 – Модель типу «чорна скринька»

Структура системи. Система «GenAI чат-бот для роботи з конфіденційною інформацією» складається з кількох ключових компонентів, які забезпе-

чують її функціональність. Основними компонентами є:

– модуль обробки запитів;

– генеративна мовна модель;

– модуль шифрування;

– хмарна інфраструктура;

– система безпеки.

Розглянемо більш детально компоненти системи.

Модуль обробки запитів виконує аналіз вхідних даних, які надходять від користувача. Він визначає, чи містить запит чутливу інформацію, та відповідно адаптує її обробку для забезпечення конфіденційності.

Генеративна мовна модель (наприклад, Claude або GPT) є центральним компонентом системи. Вона забезпечує створення відповідей, які відповідають контексту запиту. Модель адаптована для роботи із чутливою інформацією, гарантуючи її точність і відповідність заданому завданню.

Модуль шифрування забезпечує захист усіх даних, що надходять до системи та виходять з неї. Він використовує сучасні методи шифрування для гарантування конфіденційності на кожному етапі обробки.

Хмарна інфраструктура, зокрема сервіси AWS (наприклад, AWS S3 і Bedrock), відповідає за масштабованість, доступність та продуктивність системи. Вона також дозволяє зберігати тимчасові дані та забезпечує швидкий доступ до них.

Системи безпеки реалізують політики управління доступом, зокрема, через механізми AWS IAM. Вони обмежують доступ до конфіденційної інформації, базуючись на ролях користувачів, та здійснюють аудит усіх дій у системі.

Межею системи «GenAI чат-бот» є обмежене середовище, в якому забезпечується обробка запитів, зберігання даних і генерація відповідей відповідно до вимог конфіденційності. Це середовище включає в себе всі компоненти системи, які взаємодіють між собою та з зовнішніми елементами.

До основних елементів зовнішнього середовища належать:

а) користувачі – фізичні особи або організації, які надсилають запити до

системи для отримання релевантної інформації;

б) хмарні сервіси – забезпечують інфраструктурну підтримку, обробку та зберігання даних у безпечному середовищі;

в) розробники та адміністратори – відповідальні за налаштування, моніторинг і оновлення системи, включаючи політики доступу та підтримку безпеки;

г) регуляторні органи – встановлюють законодавчі та нормативні вимоги, зокрема щодо захисту даних (GDPR, HIPAA, CCPA);

д) постачальники технологій – забезпечують апаратне та програмне забезпечення для підтримки системи, включаючи сервери, шифрувальні модулі та API.

Обмін інформацією між системою та зовнішнім середовищем забезпечує ефективну роботу чат-бота, підтримує конфіденційність даних і відповідає вимогам нормативної відповідності.

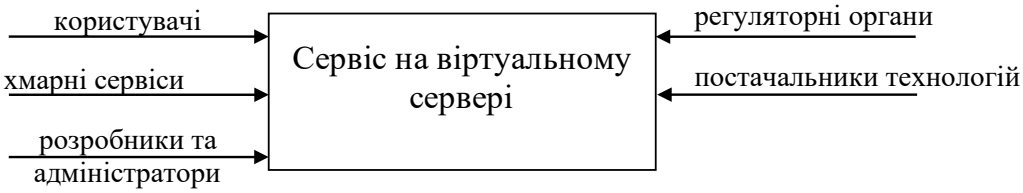

Рисунок 1.2 – Модель зовнішнього середовища системи

1.1.3 Функціональна модель системи

Метою функціональної моделі системи є аналіз процесів, які забезпечують безпечне впровадження та функціонування генеративного AI чат-бота у середовищі, що працює з конфіденційними даними. Система спрямована на забезпечення конфіденційності, надійності та відповідності нормативним вимогам.

Функціональна модель системи включає такі процеси:

– збір даних та обробка запитів: на вхід системи надходять текстові запи-

ти, що можуть містити конфіденційну інформацію (наприклад, персональні дані, фінансові звіти або медичну інформацію);

– генерація відповідей: генеративна мовна модель, наприклад Claude або GPT, використовується для створення точних та релевантних відповідей на основі контексту запиту користувача;

– шифрування даних: усі вхідні та вихідні дані проходять через модуль шифрування, який гарантує їхню безпеку під час зберігання та передачі;

– контроль доступу: політики управління ідентифікацією та доступом (IAM) забезпечують дотримання принципу мінімально необхідного доступу, лише авторизовані користувачі мають можливість взаємодіяти з системою, що забезпечує захист конфіденційної інформації.

– контроль доступу: моніторинг та оновлення системи, наприклад, за допомогою AWS CloudWatch здійснюється постійний моніторинг роботи системи для виявлення аномальної активності та разі потреби проводяться оновлення моделі для збереження її актуальності та точності;

– валідація та аудит: перед впровадженням моделі проводиться ретельна перевірка її здатності працювати з конфіденційними даними, дотримуючись стандартів захисту, логування дій забезпечує можливість аудиту та виявлення потенційних проблем.

Функціональна модель системи визначає всі ключові етапи обробки даних, від їх надходження до генерації відповіді, та забезпечує виконання вимог безпеки та нормативної відповідності. Це дозволяє інтегрувати чат-бот у середовища, які потребують суворого дотримання стандартів конфіденційності та захисту даних.

Графічне подання функціонального опису системи «Інтегрована система захищеної обробки конфіденційних запитів на основі генеративного AI» може бути виконане за допомогою контекстної діаграми IDEF0 (рис. 1.3). Входами системи є запити користувачів, які містять текстову або структуровану інформацію, що потребує обробки. Ці запити можуть включати чутливі дані, такі як персональні дані, фінансова або медична інформація.

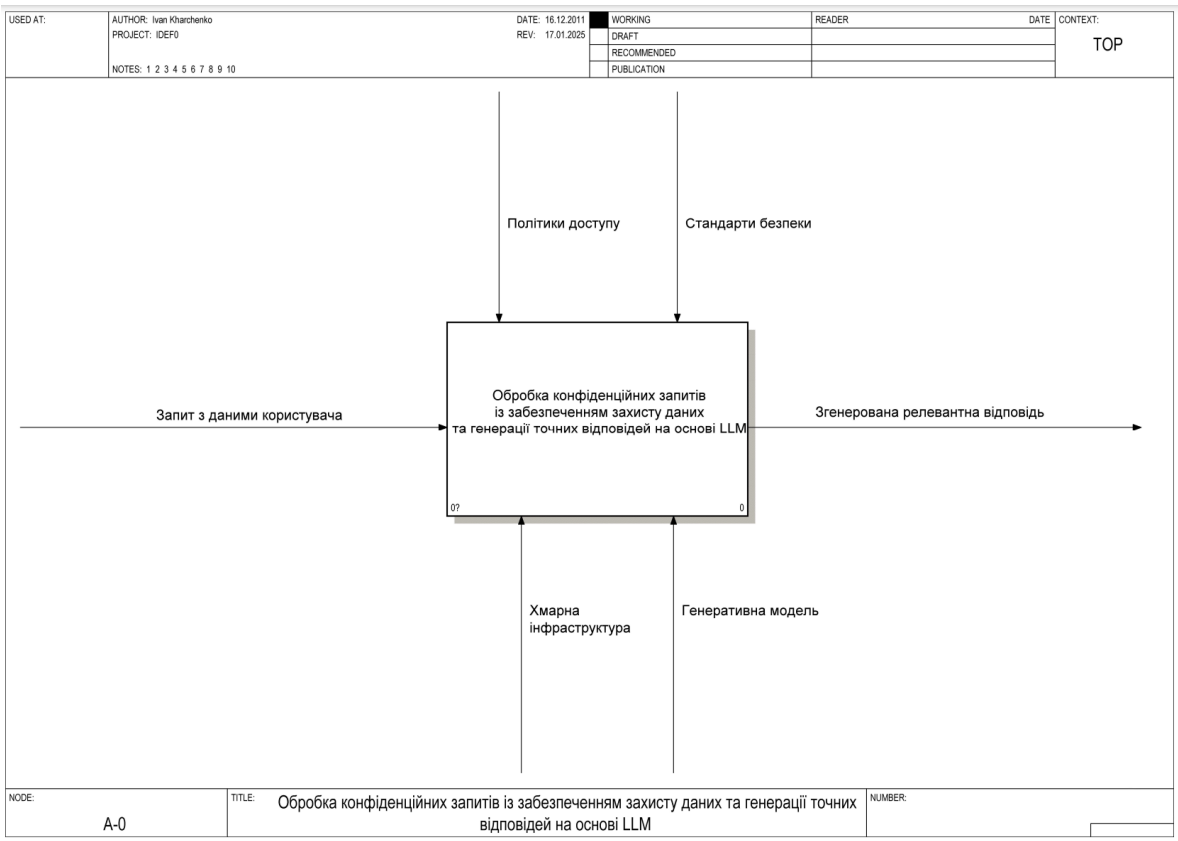

Рисунок 1.3 – Контекстна діаграма (рівень A-0)

До механізмів системи належать: генеративна мовна модель, яка забезпечує створення точних та релевантних відповідей; хмарна інфраструктура, яка підтримує масштабованість, захищене зберігання та обробку даних; модуль обробки запитів, що адаптує дані для подальшої обробки. Крім того, адміністратори системи забезпечують підтримку працездатності, налаштування доступу та моніторинг системи.

До управління системою належать правила, стандарти та стратегії, що забезпечують її функціонування. Основними управлінськими елементами є політики доступу, які обмежують доступ до системи лише авторизованим користувачам; а також стандарти безпеки, які регулюють шифрування даних та ведення журналу дій.

Виходом системи є згенеровані релевантні відповіді, які відповідають запитам користувачів і дотримуються стандартів конфіденційності та точності. Ці відповіді створюються з урахуванням контексту запитів і обробляються у захищеному середовищі.

Декомпозиція контекстної діаграми (рисунок 1.4) містить ключові етапи, необхідні для обробки запитів, а саме збір даних та ініціалізацію запиту та обробку запиту з генерацію відповіді.

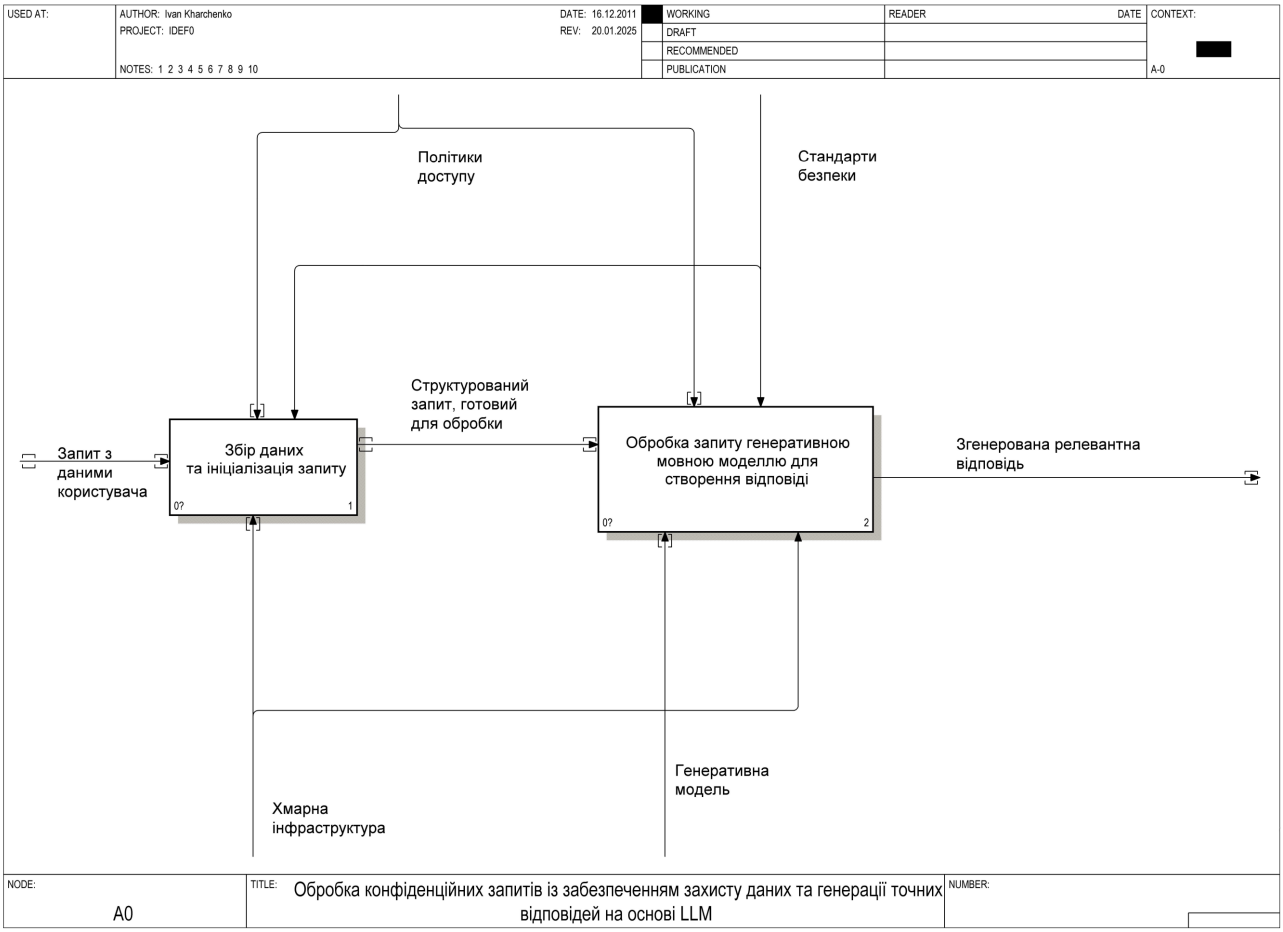

Рисунок 1.4 – Декомпозиція роботи «Інтегрована система захищеної обробки конфіденційних запитів на основі генеративного AI»: рівень A0

Система обробки конфіденційних запитів забезпечує автоматизовану роботу з даними користувача та генерує точні відповіді, дотримуючись стандартів захисту даних. Вхідним елементом системи є запит із даними користувача, який передається до модуля збору даних та ініціалізації запиту. На цьому етапі відбувається перевірка запиту на відповідність форматам, визначеним політиками доступу, і підготовка його до подальшої обробки. Результатом цього процесу є структурований запит, готовий до передачі мовній моделі.

Далі структурований запит надходить до модуля обробки, де за допомо-

гою генеративної мовної моделі, наприклад Claude чи GPT, здійснюється аналіз запиту та створення релевантної відповіді. Цей процес відбувається в середовищі, що підтримує стандарти безпеки, такі як шифрування даних і захист доступу.

Згенерована відповідь після обробки в мовній моделі повертається до користувача через захищений канал. Для забезпечення продуктивності та масштабованості всі процеси виконуються в хмарній інфраструктурі, яка гарантує надійне зберігання й передачу даних. Усі дії в системі регулюються політиками доступу, а також нормативними вимогами, які встановлюють стандарти безпеки.

1.1.4 Інформаційна модель

Інформаційна модель системи «Інтегрована система захищеної обробки запитів на основі генеративного AI» дозволяє описати структуру даних, їх взаємозв’язки та основні елементи, необхідні для забезпечення безпечної обробки конфіденційної інформації та генерації відповідей. Основними компонентами інформаційної моделі є такі:

– Дані користувача, що включають текстові або структуровані запити, які можуть містити конфіденційну інформацію, таку як персональні дані, фінансові звіти або медична інформація.

– Модуль обробки запитів, який виконує перевірку формату даних, їх верифікацію та ідентифікацію чутливих елементів.

– Генеративна мовна модель, що використовує оброблені запити для створення релевантних відповідей, забезпечуючи їх відповідність контексту та вимогам точності.

– Модуль шифрування, який гарантує захист вхідних і вихідних даних на всіх етапах обробки, використовуючи сучасні методи криптографії.

– Хмарна інфраструктура, яка забезпечує безперебійність роботи системи,

масштабованість, збереження даних та їх захищений доступ.

– Згенерована відповідь, яка є кінцевим результатом роботи системи, адаптованим до запиту користувача та готовим до передачі через захищений канал.

Графічне подання потоків даних інформаційної моделі цієї системи може бути виконане за допомогою DFD-діаграми – рисунок 1.5. Вона наочно демонструє, як дані передаються між компонентами системи, обробляються та повертаються користувачу.

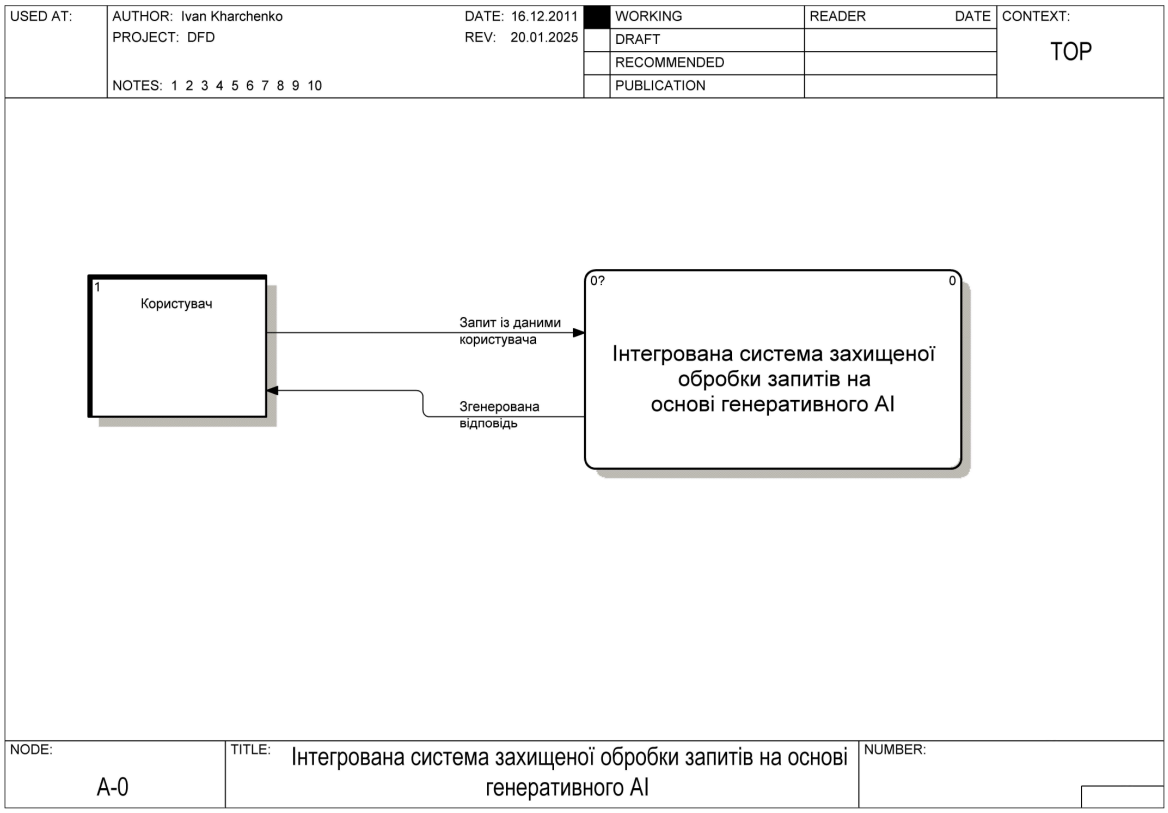

Рисунок 1.5 – Контекстна діаграма (рівень A-0)

DFD-діаграма інформаційної моделі системи подає початкову ланку потоку даних як «Користувача», який надсилає запит. Остаточним етапом є створення згенерованої відповіді, яка передається тому ж користувачу. Перший рівень декомпозиції дозволяє деталізувати ключові процеси: збір даних і ініціалізацію запиту, обробку запиту генеративною мовною моделлю, а також забезпечення безпеки даних на всіх етапах.

1.2 Аналіз сценаріїв вирішення задачі «Побудова інтегрованої системи захищеної обробки конфіденційних запитів на основі генеративного AI»

1.2.1 Модель аналізу проблеми

Наступним етапом системного аналізу задачі є вибір методу, який найбільш доцільно застосувати для створення інтегрованої системи захищеної обробки запитів із використанням генеративних мовних моделей. Для прийняття обґрунтованого рішення про вибір відповідного методу необхідно визначити перелік можливих підходів і порівняти їх за відповідними критеріями.

Для аналізу було обрано такі критерії оцінки:

– критерій 1 (К1): конфіденційність;

– критерій 2 (К2): вартість;

– критерій 3 (К3): релевантність відповідей;

– критерій 4 (К4): швидкодія.

Для реалізації поставленої задачі було визначено такі альтернативи:

– альтернатива 1 (А1): Google Gemini;

– альтернатива 2 (А2): OpenAI ChatGPT;

– альтернатива 3 (А3): власна реалізація на базі AWS Bedrock (модель Claude).

Ієрархічна модель вибору методу для вирішення задачі з побудови інтегрованої системи захищеної обробки запитів подана на рисунку 1.6. Основною метою ієрархії є забезпечення вирішення задачі, першим рівнем виступають критерії оцінки (конфіденційність, вартість, релевантність відповідей, швидкодія), а другим рівнем — альтернативи (Google Gemini, OpenAI ChatGPT, AWS Bedrock). Ця модель дозволяє порівняти альтернативи за ключовими характеристиками, обґрунтувати вибір найкращого методу та забезпечити ефективне впровадження системи.

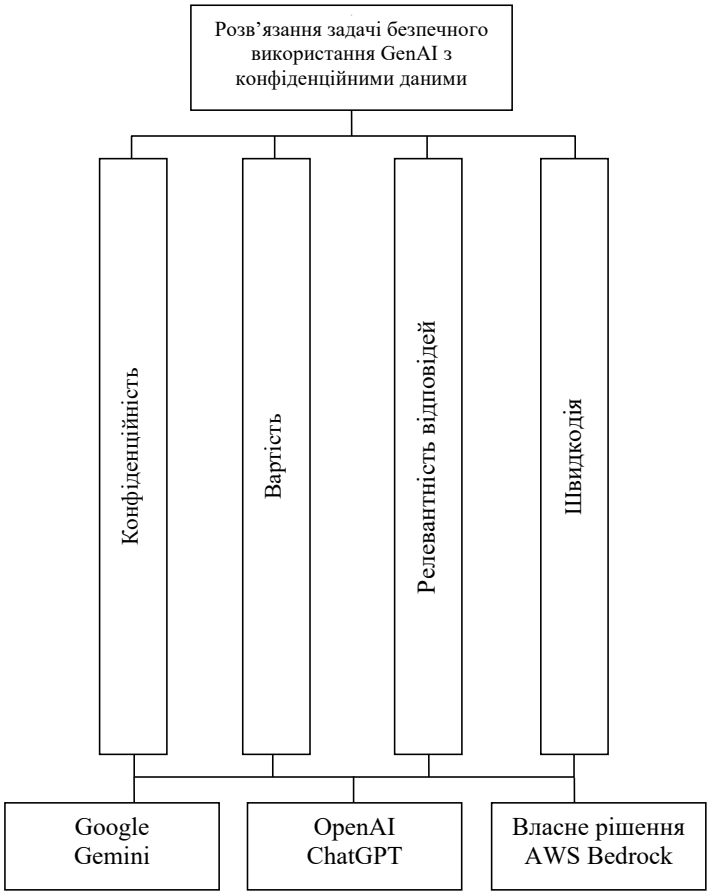

Рисунок 1.6 – Ієрархічна модель вибору методу розв’язання задачі безпечного використання GenAI з конфіденційними даними

1.2.2 Оцінювання вектора пріоритетів незадоволеностей методом аналізу ієрархій

Для аналізу ієрархії побудуємо матриці парних порівнянь моделі, а також критеріїв системи.

Матриця парних порівнянь критеріїв записана у таблиці 1.1. Останній стовпчик цієї таблиці містить результати розрахунків для вектора пріоритетів критеріїв.

Таблиця 1.1 – Матриця парних порівнянь критеріїв

Критерії оцінювання	К1	К2	К3	К4	Оцінки компонентів	Вектор пріоритетів
К1	1	5	7	9	4,21	0,65
К2	0,2	1	3	5	1,32	0,20
К3	0,14	0,33	1	3	0,61	0,10
К4	0,11	0,2	0,33	1	0,29	0,05
Усього					6,44	

Випадкова узгодженість для матриці четвертого порядку дорівнює 0,9.

За даними таблиці 1.1:

– індекс узгодженості $\text{ІУ}=\frac{4,19-4}{4-1}\approx 0,063$;

– відносна узгодженість $\text{ВУ}=\frac{0,063}{0,9}\approx 0,07=7\%$.

Оскільки відносна узгодженість близька до 0,1, то робимо висновок, що матриця парних порівнянь критеріїв побудована правильно.

Вектор локальних пріоритетів критеріїв відносно проблеми вибору дорівнює $\vec{p}^{\,K}=(0,65;0,2;0,1;0,05)$.

Сформуємо матриці попарних порівнянь альтернатив за кожним критерієм (таблиці 1.2 – 1.5) та виконаємо розрахунки за ними.

За даними таблиці 1.2:

– індекс узгодженості $\text{ІУ}=\frac{3,04-3}{3-1}\approx 0,02$;

– відносна узгодженість $\text{ВУ}=\frac{0,02}{0,58}\approx 0,03=3\%$.

Таблиця 1.2 – Матриця попарних порівнянь за першим критерієм

К1	А1	А2	А3	Оцінки компонентів	Вектор пріоритетів
А1	1	0,33	0,22	0,41	0,10
А2	3	1	0,33	1,00	0,26
А3	5	3	1	2,47	0,64
Усього				3,87	

Таблиця 1.3 – Матриця попарних порівнянь за другим критерієм

К2	А1	А2	А3	Оцінки компонентів	Вектор пріоритетів
А1	1	3	5	2,47	0,65
А2	0,33	1	2	0,87	0,23
А3	0,2	0,5	1	0,46	0,12
Усього				3,80	

Таблиця 1.4 – Матриця попарних порівнянь за третім критерієм

К3	А1	А2	А3	Оцінки ком-понентів	Вектор пріоритетів
А1	1	0,2	0,33	0,4	0,1
А2	5	1	3	2,47	0,64
А3	3,03	0,33	1	1,00	0,26
Усього				3,87	

Таблиця 1.5 – Порівняння за четвертим критерієм

К4	А1	А2	А3	Оцінки компонентів	Вектор пріоритетів
А1	1	3	0,5	1,14	0,31
А2	0,33	1	0,2	0,41	0,11
А3	2	5	1	2,15	0,58
Усього				3,7	

За даними таблиці 1.3:

– індекс узгодженості $\text{ІУ}=\frac{3,004-3}{3-1}\approx 0,002$;

– відносна узгодженість $\text{ВУ}=\frac{0,002}{0,58}\approx 0,003=0,3\%$.

За даними таблиці 1.4:

– індекс узгодженості $\text{ІУ}=\frac{3,04-3}{3-1}\approx 0,02$;

– відносна узгодженість $\text{ВУ}=\frac{0,02}{0,58}\approx 0,034=3,4\%$.

За даними таблиці 1.5:

– індекс узгодженості $\text{ІУ}=\frac{3,004-3}{3-1}\approx 0,002$;

– відносна узгодженість $\text{ВУ}=\frac{0,002}{0,58}\approx 0,003=0,3\%$.

1.2.3 Модель вирішення проблеми

Проведений аналіз у п. 1.2.2 дозволив обрати найкращий підхід для реалізації інтегрованої системи захищеної обробки конфіденційних запитів із використанням генеративного AI. Підсумкові результати оцінки альтернатив представлені в таблиці 1.6.

Таблиця 1.6 – Остаточні розрахунки

Критерій Альтернатива	К1	К2	К3	К4	Узагальнені пріоритети
А1	0,10	0,65	0,1	0,31	0,23
А2	0,26	0,23	0,64	0,11	0,28
А3	0,64	0,12	0,26	0,58	0,49

На основі отриманих даних оптимальним вибором для вирішення поставленої задачі є третя альтернатива – впровадження AWS Bedrock, яка отримала найвищий глобальний пріоритет (0.49). Таке рішення є обґрунтованим, оскільки AWS Bedrock продемонстрував найбільшу відповідність ключовим критеріям – конфіденційності (К1) та швидкодії (К4), що є основними вимогами для створення безпечної та ефективної системи.

Отже, AWS Bedrock найкраще відповідає всім вимогам і може бути рекомендований як оптимальний інструмент для побудови системи, яка забезпечує надійність, продуктивність і відповідність нормативним стандартам.

1.3 Змістовна та формальна постановка задачі

1.3.1 Змістовна постановка задачі

Обробка конфіденційних даних є критично важливим завданням для організацій, що працюють із персональною, фінансовою або медичною інформацією. Використання сучасних технологій, таких як LLM, відкриває нові можливості для автоматизації процесів обробки інформації та взаємодії з користувачами. Однак інтеграція таких моделей у реальні системи супроводжується низкою викликів, зокрема, забезпеченням безпеки та конфіденційності даних.

Поставлена задача полягає в створенні системи, яка дозволяє безпечно

обробляти конфіденційні запити, дотримуючись нормативних вимог (наприклад, GDPR, HIPAA) і забезпечуючи високий рівень точності та релевантності відповідей. Така система має бути здатною адаптуватися до різних сценаріїв використання, гарантувати швидку обробку даних і зберігати їх конфіденційність на всіх етапах роботи.

Розв'язання цієї задачі базується на застосуванні генеративних мовних моделей, інтегрованих у захищене середовище, що використовує сучасні хмарні інфраструктури, такі як AWS Bedrock. Завдяки цьому система зможе не лише автоматизувати процеси взаємодії з користувачами, але й відповідати жорстким вимогам до безпеки даних у таких сферах, як охорона здоров'я, фінанси або державне управління.

1.3.2 Формальна постановка задачі

Мета оптимізації – побудова моделі $f : X \to Y$, яка:

а) генерує відповіді $\vec{y}$, забезпечуючи високу релевантність:

$$maxP(\vec{y}|\vec{x}),$$

де $\vec{x}$ – вхідний запит,

$\vec{y}$ – відповідь;

б) забезпечує відповідність політикам безпеки S:

$$C(\vec{y},S) \le \varepsilon,$$

де $C(\vec{y},S)$ – метрика відхилення відповіді від правил безпеки,

ε – допустимий поріг.

Цільова функція – це мінімізувати втрати:

$$L(\theta) = -\frac{1}{N}\sum_{i=1}^{N} logP(y_i|\vec{x}_i) + \lambda \cdot C(\vec{y}_i, S),$$

де λ – гіперпараметр, що контролює баланс між релевантністю та безпекою.

1.4 Постановка задач дослідження

Метою кваліфікаційної роботи є створення інтегрованої системи захищеної обробки конфіденційних запитів із використанням генеративних мовних моделей. Виходячи з цього, сформулюємо перелік задач, які необхідно виконати в межах даного дослідження:

– провести огляд і аналіз сучасних підходів до побудови систем захищеної обробки даних із використанням генеративних AI-моделей;

– визначити оптимальний підхід до реалізації системи на основі порівняння існуючих рішень та вибору генеративної мовної моделі;

– розробити прототип інтегрованої системи, яка забезпечує обробку конфіденційних запитів і генерацію релевантних відповідей;

– провести тестування створеного прототипу, виконати аналіз отриманих результатів і оцінити відповідність системи вимогам до конфіденційності, релевантності та швидкодії.

Ці задачі забезпечують досягнення основної мети дослідження – створення ефективної та безпечної системи обробки даних, яка відповідає сучасним вимогам до захисту інформації та функціональності.

2 ВИБІР ТА ОБҐРУНТУВАННЯ МЕТОДУ РОЗВ'ЯЗАННЯ

2.1 Генеративні мовні моделі: поняття та сфери застосування

Генеративні мовні моделі представляють собою прогресивний напрям у сфері штучного інтелекту, що дозволяє створювати текстову інформацію на основі вхідних даних. Їх унікальна властивість полягає в здатності аналізувати контекст і генерувати семантично узгоджені відповіді, які можуть бути як точними, так і творчо адаптованими.

Фундаментом роботи генеративних мовних моделей є алгоритми глибокого навчання, натреновані на великих корпусах текстових даних. Такі моделі базуються на архітектурі трансформерів, наприклад, GPT чи BERT, що забезпечує високу якість аналізу текстових структур, розуміння контексту та генерування відповідей. Принцип роботи моделі зводиться до передбачення наступного слова або фрагмента тексту на основі попереднього контексту із застосуванням ймовірнісних методів.

Ключові характеристики генеративних мовних моделей:

– здатність генерувати текст із природним мовним стилем, адаптований до заданого контексту;

– навчання на багатомовних та багатожанрових наборах даних, що дозволяє моделі працювати у різноманітних доменах;

– гнучкість у вирішенні завдань – від автоматизації рутинних процесів до створення складних творчих текстів.

Сьогодні генеративні мовні моделі демонструють широке застосування у різних сферах, таких як:

а) обслуговування клієнтів – ці моделі активно використовуються для автоматизації відповідей у чатах, створення інтерактивних помічників і оптимізації взаємодії з клієнтами;

б) медицина – генеративні AI забезпечують створення діагностичних рекомендацій, генерацію медичних звітів та надання пацієнтам інформативних

консультацій;

в) фінансова аналітика – вони сприяють автоматизації обробки документів, створенню фінансових звітів і вдосконаленню систем навчання клієнтів щодо фінансових продуктів;

г) освіта та наукові дослідження – використовуються для генерації навчальних матеріалів, автоматичного створення тестових завдань і пояснення складних концепцій доступною мовою;

д) творчі індустрії – генеративні моделі здатні створювати художні тексти, сценарії, поезію або музичні композиції на основі заданих параметрів;

е) інформаційна безпека – вони допомагають аналізувати текстові дані для виявлення ризиків, таких як витоки інформації або порушення конфіденційності;

є) персоналізація контенту – завдяки своїй адаптивності, моделі використовуються для створення індивідуальних рекомендацій у маркетингу, електронній комерції та медіа.

Генеративні мовні моделі стали невід'ємною частиною сучасного штучного інтелекту, що стимулює значний прогрес у багатьох галузях. Їх гнучкість та функціональність відкривають нові можливості для автоматизації, творчості та оптимізації, водночас піднімаючи важливі питання безпеки, конфіденційності та етичності використання таких технологій.

2.2 Підходи до побудови захищених систем для роботи з конфіденційними даними

Обробка конфіденційної інформації потребує створення складних і багатошарових захищених систем, які відповідають сучасним вимогам безпеки, нормативним стандартам і забезпечують надійну роботу в умовах постійних загроз. Основні вимоги до таких систем зосереджуються на захисті даних під час їх зберігання, передачі й обробки, а також на ефективному управлінні досту-

пом.

Далі розглянемо ключові принципи побудови захищених систем:

Конфіденційність. Підтримка конфіденційності досягається через шифрування даних на всіх етапах їхнього життєвого циклу. Наприклад, шифрування алгоритмом AES-256 гарантує безпеку даних під час зберігання інформації, тоді як протокол TLS 1.3 забезпечує захист інформації під час передачі через мережу. Конфіденційність також вимагає обмеження доступу на основі багатофакторної автентифікації та ролей користувачів.

Цілісність даних. Цілісність інформації гарантується за допомогою хешування (наприклад, SHA-256) та цифрових підписів, що дозволяють виявляти будь-які несанкціоновані зміни в даних. Моніторинг змін у критично важливих даних забезпечує оперативне реагування на потенційні загрози, мінімізуючи ризики компрометації.

Доступність. Забезпечення доступності даних у будь-який час є фундаментальним принципом побудови захищених систем. Це досягається за допомогою резервного копіювання, реплікації даних, а також використання систем аварійного відновлення, які мінімізують час простою системи у разі збоїв. Балансування навантаження також сприяє підтриманню високої продуктивності.

Контроль доступу. Контроль доступу базується на чіткій ідентифікації користувачів, автентифікації та управлінні правами доступу. Сучасні підходи включають адаптивні системи, які оцінюють контекст доступу, наприклад геолокацію, час і тип пристрою. Принцип мінімально необхідного доступу (Least Privilege) зменшує ризик зловживання правами доступу.

Розглянемо підходи до побудови захищених систем :

– традиційні методи захисту;

– хмарні технології;

– моделі нульової довіри (Zero Trust);

– шифрування та управління ключами;

– інтеграція машинного навчання.

Традиційні методи захисту включають використання фаєрволів для об-

меження та моніторингу мережевого трафіку, що залишається ключовим елементом захисту периметра, оскільки вони фільтрують небажані запити. Важливу роль також відіграють системи виявлення та запобігання вторгнень (IDS/IPS), які аналізують активність у мережі з метою виявлення аномалій і реагують на загрози в режимі реального часу, що значно знижує вплив атак. Крім того, широко застосовується антивірусне програмне забезпечення, яке ідентифікує та нейтралізує зловмисні програми, причому регулярне оновлення сигнатур загроз є обов'язковою умовою для забезпечення ефективності захисту.

Хмарні технології забезпечують високий рівень безпеки через створення приватних хмарних інфраструктур, які ізолюють критично важливі дані, обмежуючи доступ до внутрішніх ресурсів. Інтеграція інструментів управління ключами, таких як AWS KMS, дозволяє забезпечити захищене зберігання та шифрування великих обсягів даних. Додатковий рівень захисту впроваджується через багатофакторну автентифікацію (MFA), яка використовує методи біометричної перевірки або динамічні паролі для посилення безпеки.

Моделі нульової довіри (Zero Trust) базуються на перевірці кожного запиту незалежно від його джерела, забезпечуючи автентифікацію та авторизацію всіх користувачів і пристроїв для кожної взаємодії. Важливим елементом є мікросегментація, яка розділяє мережу на ізольовані сегменти, що значно обмежує потенційний вплив атак або витоків. Крім того, постійний моніторинг поведінкових патернів користувачів дозволяє виявляти та оперативно реагувати на аномалії, підвищуючи рівень захисту.

Шифрування та управління ключами передбачають реалізацію наскрізного шифрування, яке забезпечує захист даних під час їх передачі, гарантуючи конфіденційність навіть у разі перехоплення. Шифрування на рівні транспорту додає додатковий рівень безпеки. Надійне зберігання ключів і їх використання у суворо контрольованих середовищах забезпечується за допомогою апаратних модулів безпеки (HSM).

Інтеграція машинного навчання забезпечує виявлення аномалій, які можуть свідчити про зловмисну активність, за допомогою спеціальних алгорит-

мів. Аналіз поведінкових даних дозволяє швидко ідентифікувати потенційні загрози. Прогнозні моделі сприяють передбаченню нових типів атак, що забезпечує проактивний захист системи, особливо у динамічних середовищах із високим рівнем ризику.

Створення захищених систем пов’язане з численними викликами, серед яких інтеграція багаторівневих механізмів безпеки в єдину інфраструктуру, що вимагає значних ресурсів і високого рівня технічної компетенції. Системи потребують постійної адаптації до нових загроз, які виникають через еволюцію атак і технологій, що зумовлює необхідність регулярного оновлення програмного забезпечення та інструментів. Важливим аспектом є пошук балансу між рівнем безпеки та продуктивністю системи, оскільки надмірна захищеність може негативно впливати на ефективність операцій. Додатково, значними залишаються витрати на підтримку інфраструктури безпеки, включаючи апаратні й програмні засоби, а також навчання персоналу.

Побудова захищених систем для роботи з конфіденційною інформацією вимагає інтеграції сучасних технологій, таких як моделі нульової довіри, машинне навчання та хмарні рішення, з традиційними підходами до захисту. Ця синергія дозволяє створити адаптивну та масштабовану інфраструктуру, здатну забезпечити безпеку даних у динамічному інформаційному середовищі. Майбутні дослідження повинні зосереджуватися на розробці ефективних і стійких до атак рішень, які відповідають викликам сучасного кіберпростору.

2.3 Вибір генеративної мовної моделі для інтегрованої системи

2.3.1 Використання Google Gemini у системі обробки конфіденційних запитів

Google Gemini є провідною генеративною мовною моделлю, інтегрованою в екосистему Google Cloud, яка представляє собою новітній підхід до ав-

томатизації роботи з конфіденційними запитами. Завдяки своїм унікальним можливостям, таким як потужна архітектура, багатомовність і здатність до адаптації, Google Gemini відкриває нові перспективи для організацій, що працюють із великими обсягами текстових даних. Проте, впровадження цієї моделі вимагає врахування технічних, організаційних і фінансових аспектів.

У основі Google Gemini лежить вдосконалена архітектура PaLM (Pathways Language Model), яка використовує передові трансформерні алгоритми. Ця модель створена для ефективної обробки текстових наборів даних великого обсягу, що охоплюють численні домени та мови. Завдяки своїй багатомовності Google Gemini дозволяє забезпечити високу якість відповіді навіть у складних контекстах, таких як технічна документація, фінансові звіти чи медичні записи. Її здатність адаптуватися до багатомовного середовища робить її незамінним інструментом для глобальних організацій.

Інтеграція з Google Cloud забезпечує моделі доступ до передових технологій захисту даних. Хмарна інфраструктура Google підтримує захищені протоколи передачі даних, такі як TLS 1.3, та інструменти для управління ключами й контролю доступу. Це дозволяє Google Gemini відповідати сучасним нормативним стандартам, зокрема GDPR і HIPAA. Проте централізований характер обробки даних може створювати виклики для галузей, що потребують суворого контролю над конфіденційною інформацією. Для подолання цих обмежень Google працює над розробкою функцій, які дадуть змогу доопрацьовувати модель на основі закритих наборів даних, що відкриє нові можливості для вузькоспеціалізованих завдань.

Вартість використання Google Gemini базується на обсягах оброблених даних і споживаних обчислювальних ресурсах. Google пропонує тарифні плани, орієнтовані на корпоративних користувачів, які адаптуються до їхніх потреб. Проте, інтеграція цієї моделі може потребувати значних інвестицій, включаючи оновлення інфраструктури Google Cloud та налаштування системи. Для багатьох організацій важливо враховувати ці аспекти при оцінці економічної доцільності впровадження моделі.

Швидкодія є однією з ключових переваг Google Gemini. Оптимізована архітектура моделі дозволяє забезпечити обробку запитів у реальному часі, що робить її ідеальним рішенням для інтерактивних застосувань, таких як чат-боти, автоматизовані платформи підтримки клієнтів або аналітичні інструменти. Завдяки своїй швидкості Google Gemini ефективно працює у середовищах, де оперативність є важливою умовою.

Однак використання Google Gemini також супроводжується певними викликами. Одним із основних обмежень є недостатня гнучкість у кастомізації для специфічних бізнес-потреб. Крім того, централізована обробка даних може викликати занепокоєння у галузях, де конфіденційність даних має критичне значення. Щоб вирішити ці проблеми, організації можуть розглядати гібридні моделі впровадження, які поєднують локальну обробку даних із використанням хмарної інфраструктури.

Google продовжує вдосконалювати модель Gemini, додаючи нові функції для розширення її можливостей. Серед перспективних напрямів розвитку – підтримка навчання на закритих наборах даних, інтеграція із системами автоматизації та покращення адаптації для вузькоспеціалізованих завдань. Це робить Google Gemini універсальним інструментом, придатним для використання у різних галузях, включаючи право, науку, освіту чи високотехнологічне виробництво.

Таким чином, Google Gemini є потужним рішенням для обробки конфіденційних запитів, яке поєднує багатомовність, швидкодію та точність. Її інтеграція з Google Cloud забезпечує високий рівень безпеки даних і масштабованості, що робить її оптимальним вибором для організацій, які прагнуть поєднати інновації з ефективністю. Проте, впровадження цієї моделі потребує детального аналізу витрат і специфіки бізнес-процесів, що дозволить максимально реалізувати її потенціал у відповідності до потреб організації.

2.3.2 Використання OpenAI GPT у системі обробки конфіденційних запитів

OpenAI GPT є передовою генеративною мовною моделлю, яка відіграє ключову роль у сучасних інформаційних системах, орієнтованих на автоматизацію обробки текстових даних. Її здатність до контекстуального аналізу, масштабованість та гнучкість інтеграції роблять її універсальним інструментом для вирішення складних завдань. Однак впровадження цієї моделі у системи обробки конфіденційних запитів вимагає детального врахування її технічних характеристик, фінансових аспектів і викликів, пов'язаних із безпекою даних.

Основою OpenAI GPT є трансформерна архітектура, що дозволяє моделі обробляти великі обсяги тексту, забезпечуючи генерацію логічно узгоджених і точних відповідей. Найновіші версії, такі як GPT-3.5, GPT-4 та GPT-4 Turbo, пропонують значні покращення у здатності моделі до багатомовної обробки тексту, роботи з багаторівневими діалогами та адаптації до вузькоспеціалізованих сценаріїв. GPT-4, зокрема, вирізняється своєю здатністю обробляти складні технічні документи та генерувати код. Це робить її ідеальним вибором для галузей, де точність і контекстуальність є критичними, таких як медицина, право, освіта та фінанси.

Інтеграція OpenAI GPT у корпоративні системи є відносно простою завдяки доступу до добре задокументованого API. Ця модель дозволяє організаціям налаштовувати параметри генерації, створювати спеціалізовані інтерфейси та розробляти рішення для багатомовних середовищ. Однак основна обробка даних відбувається у хмарному середовищі OpenAI, що вимагає від організацій впровадження додаткових заходів безпеки, таких як шифрування перед передачею даних і контроль доступу. Забезпечення відповідності нормативним стандартам, включаючи GDPR і HIPAA, є обов'язковим при роботі з конфіденційною інформацією.

Цінова політика OpenAI GPT базується на кількості оброблених токенів, що надає організаціям гнучкість у виборі тарифних планів відповідно до їхніх

бюджетів і потреб. GPT-4 Turbo є більш економічним рішенням для великих обсягів запитів, тоді як GPT-4 забезпечує максимальну точність і контекстуальність для завдань, що вимагають складного аналізу. Проте впровадження GPT у системи з високим навантаженням може бути фінансово витратним, тому організації повинні ретельно оцінювати доцільність її використання з урахуванням очікуваних результатів і витрат.

Однією з ключових переваг OpenAI GPT є її здатність генерувати відповіді з високою релевантністю. Завдяки навчанню на масштабних і різнорідних наборах даних, модель демонструє високу адаптивність до специфічних вимог користувачів. Це робить її надзвичайно корисною для автоматизації завдань у таких сферах, як підтримка клієнтів, створення технічної документації, аналіз даних і освітні програми. Її здатність до налаштування дозволяє забезпечити ефективну роботу навіть у найскладніших сценаріях.

Однак впровадження OpenAI GPT супроводжується певними викликами. Одним із найбільш значущих є забезпечення конфіденційності даних, оскільки обробка запитів здійснюється на стороні OpenAI. Це створює потенційні ризики, які можна мінімізувати шляхом впровадження протоколів шифрування, регулярного аудиту системи та впровадження суворого контролю доступу. У таких галузях, як медицина чи банківська справа, забезпечення відповідності нормативним вимогам є критично важливим для збереження довіри клієнтів і дотримання законодавства.

OpenAI постійно вдосконалює свої моделі, додаючи нові функції для вирішення складніших завдань. GPT-4, наприклад, підтримує створення коду, проведення багатоступеневих аналітичних оглядів та генерування тексту з високим рівнем деталізації. Це розширює можливості її використання у таких сферах, як наукові дослідження, освітні програми, юридичні консультації та творчі індустрії. Інноваційність цих рішень дозволяє організаціям використовувати GPT у дедалі складніших сценаріях, підвищуючи продуктивність і точність систем.

Таким чином, OpenAI GPT є одним із найефективніших інструментів для

обробки конфіденційних запитів, поєднуючи високу точність, адаптивність і масштабованість. Проте її впровадження вимагає ретельного аналізу викликів, пов'язаних із конфіденційністю та фінансовими витратами. Для завдань, які потребують високої якості, релевантності та здатності працювати з великими обсягами даних, GPT залишається одним із провідних рішень у галузі генеративних мовних моделей.

2.3.3 Використання AWS Bedrock (Claude) у системі обробки конфіденційних запитів

AWS Bedrock, що інтегрує генеративну мовну модель Claude від Anthropic, пропонує інноваційні рішення для систем, орієнтованих на обробку конфіденційних запитів. Завдяки поєднанню передових технологій машинного навчання, масштабованої хмарної інфраструктури та високих стандартів безпеки, ця платформа є одним із провідних інструментів для організацій, які працюють із чутливими даними.

Claude, як складова AWS Bedrock, є передовою мовною моделлю, розробленою з акцентом на етичне використання та безпеку даних. Її архітектура враховує суворі принципи конфіденційності та відповідає міжнародним стандартам, включаючи GDPR і HIPAA. Версія Claude 3.5 Sonnet демонструє виняткову точність генерації тексту, релевантність відповідей і можливість адаптації через навчання на приватних наборах даних. Це дозволяє організаціям адаптувати модель до специфічних потреб, забезпечуючи точність і ефективність її роботи навіть у складних сценаріях.

Інтеграція AWS Bedrock у систему обробки конфіденційних запитів надає доступ до широкого спектру сервісів Amazon. Основними компонентами є AWS Identity and Access Management , який забезпечує гнучке управління доступом, AWS Key Management Service для надійного шифрування даних і AWS CloudWatch для моніторингу системи. Ця багаторівнева інфраструктура дозво-

ляє захищати інформацію на кожному етапі її обробки – від прийому запиту до передачі результату. Додатково AWS PrivateLink забезпечує ізоляцію даних від загальнодоступних мереж, що значно знижує ризик витоку інформації.

Однією з головних переваг AWS Bedrock є її масштабованість. Хмарна інфраструктура Amazon дозволяє динамічно адаптувати ресурси до потреб організації, підтримуючи стабільну продуктивність навіть за умов пікових навантажень. Це особливо важливо для великих організацій, які обробляють значні обсяги даних або надають послуги у режимі реального часу. Гнучкість у налаштуванні та автоматизоване управління ресурсами дозволяють організаціям зберігати високу якість обробки даних, адаптуючись до змін у робочих процесах.

Фінансова ефективність AWS Bedrock також заслуговує на увагу. Витрати на використання моделі базуються на обсягах оброблених даних і використаних обчислювальних ресурсах. Незважаючи на початкові інвестиції в адаптацію інфраструктури та навчання персоналу, довгострокова економічна ефективність забезпечується завдяки можливостям оптимізації витрат. AWS надає інструменти для моніторингу витрат, що дозволяє організаціям ретельно планувати бюджети та контролювати фінансові витрати.

Claude у складі AWS Bedrock забезпечує високу релевантність відповідей завдяки своєму навчальному процесу, який орієнтований на етичні принципи та забезпечення конфіденційності. Модель може обробляти складні запити з винятковою точністю, що робить її ідеальною для галузей, де точність є критично важливою. Завдяки підтримці доопрацювання на приватних наборах даних, Claude дозволяє адаптувати свої алгоритми до специфічних бізнес-вимог, забезпечуючи відповідність результатів очікуванням користувачів.

Незважаючи на переваги, впровадження AWS Bedrock може супроводжуватися певними викликами. По-перше, інтеграція вимагає значних ресурсів і технічної експертизи. По-друге, організації повинні ретельно планувати витрати, особливо якщо вони працюють із великими обсягами даних. Проте ці виклики компенсуються високою продуктивністю, масштабованістю та рівнем

безпеки, які забезпечує платформа.

AWS Bedrock також пропонує розширені можливості кастомізації Claude. Наприклад, організації можуть навчати модель на закритих наборах даних, що дозволяє створювати рішення, оптимізовані для конкретних бізнес-процесів. Це особливо актуально для фінансового аналізу, медичних досліджень або правозастосування, де потрібна висока персоналізація рішень. Крім того, AWS постійно впроваджує нові функції, спрямовані на підвищення продуктивності та зниження упередженості у відповідях моделі.

Таким чином, AWS Bedrock (Claude) є комплексним рішенням для систем обробки конфіденційних запитів. Його можливості з кастомізації, навчання на специфічних даних і масштабування, у поєднанні з потужними засобами безпеки, роблять цю платформу одним із найбільш ефективних інструментів для організацій, які прагнуть забезпечити високий рівень захисту даних і ефективність роботи своїх інформаційних систем.

2.4 Обґрунтування вибору AWS Bedrock для реалізації системи

Вибір AWS Bedrock для реалізації системи обробки конфіденційних запитів є стратегічно обґрунтованим через її інноваційні технології, здатність до масштабування, відповідність міжнародним стандартам безпеки та інтеграційну гнучкість. Ця платформа поєднує передові можливості хмарної інфраструктури з високим рівнем безпеки й етичними принципами роботи з даними, що робить її універсальним інструментом для сучасних організацій.

Конфіденційність і безпека. AWS Bedrock забезпечує багаторівневий захист даних, застосовуючи провідні технології контролю доступу та шифрування. Інтеграція AWS Identity and Access Management дозволяє налаштовувати детальні політики доступу, які відповідають принципу мінімального привілею, а AWS Key Management Service гарантує використання передових алгоритмів шифрування для забезпечення безпеки конфіденційної інформації. AWS

PrivateLink створює додатковий рівень захисту, ізолюючи дані від публічних мереж і мінімізуючи ризик їх витоку. Claude, інтегрована у AWS Bedrock, підтримує навчання на закритих наборах даних, що додає додатковий рівень контролю та конфіденційності. Ця особливість дозволяє організаціям адаптувати модель до своїх бізнес-потреб, водночас зберігаючи безпеку даних на всіх етапах обробки. Для секторів, де захист даних є критичним, таких як медицина, фінанси або державне управління, AWS Bedrock пропонує незамінний набір інструментів.

Масштабованість і продуктивність. Динамічна масштабованість AWS Bedrock є однією з її найбільш значущих переваг. Хмарна інфраструктура Amazon забезпечує автоматичне розподілення ресурсів залежно від обсягу запитів, підтримуючи стабільну продуктивність навіть за умов пікових навантажень. Ця особливість є важливою для великих корпорацій, які працюють із великими обсягами даних або забезпечують безперервну обробку запитів у режимі реального часу. Інтеграція AWS Bedrock із такими сервісами, як AWS CloudWatch, додає можливості для моніторингу продуктивності системи, що дозволяє виявляти проблеми та вирішувати їх у реальному часі. Це сприяє мінімізації простоїв і забезпечує високий рівень надійності системи. Додаткові можливості, як-от використання AWS Lambda для автоматизації процесів, роблять платформу ще більш ефективною у забезпеченні продуктивності.

Економічна ефективність. Хоча інтеграція AWS Bedrock може потребувати значних початкових інвестицій, її довгострокова економічна ефективність забезпечується завдяки оптимальному використанню ресурсів. Гнучкі моделі ціноутворення, засновані на обсягах оброблених даних, дозволяють організаціям адаптувати витрати до своїх бізнес-цілей. AWS також пропонує інструменти для прогнозування витрат і моніторингу фінансових ресурсів, що допомагає зменшити перевитрати й ефективно управляти бюджетами.

Інноваційність та адаптивність. Claude, інтегрована у AWS Bedrock, забезпечує високий рівень адаптивності завдяки можливості навчання на специфічних наборах даних. Це дозволяє організаціям розробляти кастомізовані рі-

шення, оптимізовані для їхніх унікальних бізнес-процесів. AWS постійно вдосконалює функціонал платформи, впроваджуючи інноваційні функції, що покращують продуктивність і забезпечують відповідність сучасним вимогам автоматизації процесів. Додатково, платформа сприяє створенню комплексних екосистем для управління бізнес-процесами через інтеграцію з іншими інструментами AWS. Це спрощує автоматизацію, знижує операційні витрати й підвищує ефективність робочих процесів. Інноваційний підхід до зменшення упередженості в алгоритмах Claude забезпечує етичність використання генеративних мовних моделей.

Глобальна відповідність нормативним вимогам. AWS Bedrock відповідає стандартам GDPR, HIPAA, CCPA та інших регуляторних норм, що дозволяє її використовувати у різних юрисдикціях. Відповідність цим нормативам гарантує, що платформа підходить для багатонаціональних організацій, які працюють у складних регуляторних умовах. Крім того, AWS Bedrock підтримує прозорість у роботі з даними, що сприяє підвищенню довіри користувачів.

AWS Bedrock є потужним, універсальним і надійним рішенням для реалізації систем обробки конфіденційних запитів. Її здатність до масштабування, інтеграція передових технологій безпеки, економічна ефективність та інноваційність роблять її ідеальним вибором для організацій, які прагнуть забезпечити високу продуктивність і відповідність нормативним вимогам. Claude додає платформі гнучкість і функціональність, дозволяючи адаптувати її до специфічних потреб бізнесу. Таким чином, AWS Bedrock не лише відповідає сучасним викликам, але й забезпечує основу для подальших інновацій у сфері обробки даних і автоматизації.

Висновки за розділом 2

Другий розділ цієї роботи присвячено ретельному аналізу сучасних генеративних мовних моделей та обґрунтуванню вибору AWS Bedrock (Claude) як

платформи для створення системи обробки конфіденційних запитів. Результати проведеного дослідження дозволяють сформулювати важливі висновки.

Генеративні мовні моделі є фундаментом для автоматизації роботи з конфіденційними запитами. Аналіз розглянутих моделей — Google Gemini, OpenAI GPT і Claude — продемонстрував їхню високу точність, релевантність відповідей та здатність адаптуватися до різноманітних завдань. Водночас ефективність їхнього використання залежить від рівня інтеграції, відповідності вимогам безпеки та фінансових витрат.

Claude у складі AWS Bedrock вирізняється високим рівнем безпеки, масштабованістю та гнучкістю в налаштуванні. Можливість навчання моделі на закритих наборах даних дозволяє адаптувати алгоритми до унікальних потреб організації, забезпечуючи високу точність і безпеку обробки інформації. Це є особливо важливим для галузей із суворими вимогами до конфіденційності, таких як охорона здоров’я, фінансові послуги та державне управління.

AWS Bedrock надає комплексний набір інструментів для інтеграції та підтримки системи. Зокрема, сервіси AWS Identity and Access Management, AWS Key Management Service та AWS CloudWatch забезпечують багаторівневу безпеку, контроль доступу та можливості моніторингу. Такі функції мінімізують ризики витоків даних і підвищують стабільність функціонування системи.

Відповідність міжнародним стандартам безпеки є важливою перевагою платформи. AWS Bedrock відповідає вимогам GDPR, HIPAA та інших регуляторних норм, що дозволяє її використання у різних галузях і географічних регіонах. Така відповідність є критичною для організацій, які працюють у регульованих середовищах.

Економічна ефективність AWS Bedrock підкреслює її конкурентоспроможність. Гнучка модель ціноутворення, яка враховує обсяги оброблених даних, у поєднанні з інструментами для управління витратами, дозволяє організаціям оптимізувати свої фінансові ресурси. Це робить платформу привабливою як для малих і середніх підприємств, так і для великих корпорацій.

На підставі аналізу можна зробити висновок, що AWS Bedrock є оптима-

льною платформою для реалізації інтегрованої системи обробки конфіденційних запитів. Високий рівень безпеки, можливості масштабування, гнучкість у налаштуванні та економічна ефективність роблять її незамінною для сучасних організацій, які прагнуть досягти високої продуктивності та відповідності нормативним вимогам. AWS Bedrock також забезпечує перспективи для впровадження інновацій, що відкриває нові горизонти для автоматизації та оптимізації бізнес-процесів.

3 ПРОГРАМНА РЕАЛІЗАЦІЯ

3.1 Terraform як інструмент для автоматизації створення інфраструктури в AWS

Terraform є висококласним засобом управління інфраструктурою як кодом (IaC), який надає можливість автоматизувати створення, модифікацію та видалення інфраструктурних ресурсів у провідних хмарних середовищах, включаючи AWS. Його використання суттєво підвищує ефективність управління, забезпечуючи повторюваність, масштабованість і прозорість, що є ключовими вимогами при створенні захищеного середовища для роботи з конфіденційними даними.

Terraform демонструє низку функціональних можливостей, які роблять його незамінним для автоматизації хмарної інфраструктури:

а) HCL (HashiCorp Configuration Language): декларативна мова конфігурації, що вирізняється зрозумілістю і структурованістю, дозволяючи описувати ресурси та їхні залежності у формі коду;

б) мультихмарність: підтримка численних провайдерів, таких як AWS, Microsoft Azure і Google Cloud, забезпечує широкі можливості для інтеграції та управління інфраструктурою у гібридних середовищах;

в) стан інфраструктури: Terraform зберігає поточний стан ресурсів у спеціальному state-файлі, що дозволяє ефективно синхронізувати бажаний і фактичний стан системи;

г) модульність: завдяки модулям, Terraform сприяє створенню повторно використовуваного і легко підтримуваного коду, що зменшує ризик помилок;

д) інтеграція з AWS: можливість використовувати IAM для впровадження політик мінімального доступу є критично важливою для забезпечення безпеки;

е) масштабованість: паралельне створення ресурсів знижує час розгортання великих інфраструктур.

Terraform значно полегшує процес автоматизації складних архітектур. Він

спрощує розгортання таких елементів, як VPC, багаторівневі підмережі, зашифровані S3-бакети та системи моніторингу. Це дозволяє ефективно впроваджувати складні рішення без значних ручних втручань.

Однією з головних переваг є можливість реплікації середовищ. Завдяки кодовій базі Terraform, легко створюються середовища розробки, тестування та продакшну з однаковими характеристиками, що сприяє стандартизації та стабільності.

Прозорість є ще однією важливою рисою Terraform. Планування змін перед їх застосуванням дозволяє чітко розуміти вплив кожної операції, забезпечуючи контроль над інфраструктурою.

Моніторинг стану є ключовою функцією. Завдяки збереженню state-файлів Terraform дозволяє аналізувати конфігураційні відмінності між середовищами, що значно підвищує їхню керованість.

Інтеграція з AWS-сервісами робить Terraform універсальним інструментом. Підтримка ключових сервісів AWS, таких як EC2, RDS, S3, Lambda, забезпечує точне налаштування ресурсів відповідно до вимог безпеки та продуктивності.

Terraform є потужним інструментом для створення захищених архітектур, зокрема для GenAI чат-ботів. Його використання дозволяє забезпечити ізоляцію інфраструктури через розгортання приватних мереж (VPC), налаштування підмереж та відповідних маршрутів для мінімізації ризиків витоку даних.

Автоматизація створення зашифрованих S3-бакетів із політиками шифрування на рівні даних гарантує безпечне зберігання інформації, що є критично важливим для чутливих даних.

Контроль доступу також досягається завдяки використанню IAM-ролей, які забезпечують мінімально необхідні привілеї та підтримують принцип розподілу прав доступу.

Розгортання приватних кінцевих точок за допомогою AWS PrivateLink дозволяє ізолювати трафік від публічних мереж, підвищуючи рівень захищеності.

Terraform забезпечує впровадження архітектур, які відповідають високим стандартам інформаційної безпеки та регуляторним вимогам, таким як GDPR і HIPAA. Декларативний підхід до конфігурацій та підтримка інтеграції з провідними сервісами AWS дозволяють досягати передбачуваного, надійного і швидкого розгортання, що особливо важливо для проєктів зі строгими вимогами до безпеки.

3.2 Алгоритм розв'язання задачі створення захищеного середовища для GenAI чат-бота

Розробка захищеного середовища для GenAI чат-бота є комплексним завданням, яке вимагає багатоетапного підходу, що забезпечує не лише функціональність, але й повну відповідність стандартам безпеки та нормативним вимогам. Нижче детально описано ключові етапи, які реалізують створення такої інфраструктури.

Крок 1. Налаштування VPC, підмереж і груп безпеки.

Створення віртуальної приватної мережі (VPC) забезпечує ізоляцію трафіку між компонентами системи. Це включає розподіл CIDR-блоків, враховуючи майбутнє масштабування, та використання інструментів Terraform для автоматизації й документування VPC-конфігурації. Під час налаштування підмереж приватні підмережі виділяються для ресурсів із чутливими даними, обмежуючи їхній доступ ззовні, а публічні підмережі використовуються для ресурсів із потребою виходу в Інтернет із застосуванням NAT. Управління трафіком між підмережами та зовнішнім середовищем здійснюється через таблиці маршрутизації.

Групи безпеки налаштовуються для контролю вхідного та вихідного трафіку, дозволяючи доступ лише до необхідних портів і IP-адрес. Для додаткового контролю впроваджуються ACL. Регулярний аудит конфігурацій забезпечує відповідність політикам безпеки та захист від потенційних ризиків.

Крок 2. Використання AWS S3 для зберігання даних із шифруванням.

Для ефективного зберігання даних створюються S3-бакети для різних категорій, таких як журнали, файли моделей і тренувальні дані. Визначаються Lifecycle-політики для автоматичного управління архівуванням і видаленням старих даних. Шифрування даних реалізується через серверне шифрування (SSE) з AWS KMS, а передача даних захищається протоколами SSL/TLS. Регулярне тестування гарантує цілісність механізмів шифрування.

Політики доступу налаштовуються через IAM для обмеження доступу залежно від ролей і привілеїв користувачів. Для захисту від випадкового видалення чи перезапису застосовується контроль версійності.

Крок 3. Розгортання приватної кінцевої точки (AWS PrivateLink).

PrivateLink дозволяє використовувати приватні кінцеві точки для доступу до критичних сервісів AWS без виходу в Інтернет. Налаштування інтегрується із VPC Endpoint Policies для обмеження доступу за ролями. Моніторинг трафіку через VPC Flow Logs допомагає виявляти аномальну активність, а доступ до кінцевих точок обмежується конкретними IP-адресами або підмережами.

Для забезпечення прозорості налаштовуються AWS CloudWatch і CloudTrail. CloudWatch відстежує активність та ідентифікує підозрілі дії, а CloudTrail веде записи всіх взаємодій з кінцевими точками, дозволяючи проводити детальні аудити.

Крок 4. Конфігурація IAM для обмеження доступу.

Розробляються ролі IAM із принципом «найменш необхідних привілеїв», забезпечуючи мінімальний рівень доступу для виконання завдань. Використання умов (Condition Keys), таких як час доби чи геолокація, підвищує безпеку доступу.

MFA впроваджується для критичних ролей і облікових записів із доступом до конфіденційної інформації. Використовуються фізичні ключі безпеки як додатковий рівень захисту. Регулярний аудит дозволів за допомогою AWS Access Analyzer виявляє надмірні привілеї, а механізми автоматичного відкликання доступу відключають користувачів, які тривалий час не взаємодіяли з си-

стемою.

Виконання описаних етапів забезпечує створення інфраструктури, яка не лише відповідає сучасним стандартам безпеки, але й легко адаптується до змінних потреб організації. Комплексний підхід до управління VPC, S3, PrivateLink та IAM дозволяє мінімізувати ризики витоку інформації та забезпечити конфіденційність даних на кожному етапі їх обробки. У результаті досягається не лише функціональність, але й надійність системи, що критично важливо для розгортання GenAI чат-бота в середовищі з підвищеними вимогами до інформаційної безпеки.

3.3 Опис програми

У цьому розділі представлено детальний опис написаного Terraform-коду, який забезпечує створення необхідної інфраструктури для GenAI чат-бота, а також приклади основних конфігураційних файлів.

Основні файли Terraform:

а) файл main.tf – цей файл містить основну логіку проєкту та взаємозв’язки між усіма компонентами інфраструктури. Він включає:

1) виклик модулів (VPC, S3, Bedrock і IAM);

2) налаштування провайдерів (AWS);

3) основні ресурси, які не входять до модулів, але необхідні для зв’язку компонентів;

б) файл variables.tf – містить усі змінні, які використовуються в конфігураціях. Приклади:

1) регіон AWS;

2) ідентифікатори ресурсів;

3) параметри для налаштувань моделі (наприклад, ім’я S3-бакету, ID моделі);

в) файл outputs.tf – використовується для визначення результатів, які мо-

жуть бути корисними після виконання Terraform. Приклади вихідних даних:

1) URL-адреса кінцевої точки Bedrock;

2) ARN створених IAM ролей;

3) ID VPC і підмереж.

Використовувані ресурси Terraform:

а) мережеві ресурси (VPC):

1) VPC – ізольована мережа для розгортання ресурсів;

2) Public і Private Subnets – публічні та приватні підмережі для різних типів ресурсів;

3) NAT Gateway – забезпечує приватним ресурсам доступ до Інтернету;

4) Route Tables – таблиці маршрутизації для керування трафіком між підмережами;

б) сховище даних (S3):

1) S3-бакети – створення сховищ для зберігання даних;

2) шифрування – автоматичне використання шифрування за допомогою AWS KMS для захисту даних;

3) політики доступу – IAM-політики для обмеження доступу;

в) інтеграція з AWS Bedrock:

1) PrivateLink Endpoint – приватна кінцева точка для доступу до API Bedrock;

2) модель Claude – використовується для обробки запитів і надання відповідей;

3) знання (Knowledge Base) – інтеграція з S3 для завантаження даних моделі;

г) ролі та політики доступу (IAM):

1) IAM Roles – створення ролей для сервісів, які взаємодіють із ресурсами.

2) IAM Policies – політики доступу для забезпечення принципу мінімально необхідних привілеїв;

3) мультифакторна автентифікація (MFA) – впровадження додаткового рівня безпеки;

д) інші ресурси:

1) CloudWatch – для моніторингу і ведення журналів діяльності;

2) Outputs – автоматичне отримання важливих значень (наприклад, URL-адреси).

Опис ключових компонентів:

а) модуль VPC – забезпечує створення ізольованого середовища з приватними та публічними підмережами, NAT-шлюзами та таблицями маршрутизації;

б) модуль S3 – відповідає за створення захищеного сховища даних з шифруванням та контрольованим доступом;

в) модуль Bedrock – налаштовує інтеграцію з AWS Bedrock, включаючи доступ до моделі Claude через приватну кінцеву точку;

г) модуль IAM – забезпечує налаштування ролей і політик доступу з використанням принципу мінімально необхідних привілеїв;

д) додаткові сервіси – включають моніторинг через CloudWatch для забезпечення стабільної роботи інфраструктури.

Висновки за розділом 3

У третьому розділі було проведено детальне дослідження програмної реалізації інфраструктури для GenAI чат-бота в захищеному середовищі на основі AWS.

У підрозділі 3.1 було обґрунтовано вибір Terraform як інструмента для автоматизації створення інфраструктури. Розглянуто його ключові переваги, зокрема декларативний підхід, модульність, повторюваність конфігурацій і підтримка інтеграції з багатьма хмарними сервісами. Відзначено, що Terraform дозволяє ефективно керувати складною архітектурою з урахуванням вимог інфо-

рмаційної безпеки.

У підрозділі 3.2 описано алгоритм створення захищеного середовища. Включено кроки, пов'язані з налаштуванням мережевої інфраструктури (VPC, підмережі, NAT-шлюзи), шифрованого сховища даних (S3), приватної кінцевої точки (AWS PrivateLink) і ролей доступу (IAM). Представлений алгоритм забезпечує відповідність сучасним стандартам безпеки та ефективності.

У підрозділі 3.3 детально описано реалізацію коду Terraform, який відповідає за автоматизоване розгортання інфраструктури. Виділено основні компоненти конфігураційних файлів, зокрема main.tf, variables.tf і outputs.tf. Підкреслено, що модульний підхід дозволяє легко масштабувати проєкт, адаптувати його до нових вимог і забезпечує надійне управління інфраструктурою.

Таким чином, програмна реалізація за допомогою Terraform дозволяє створити надійне та захищене середовище для роботи GenAI чат-бота. Модульний підхід забезпечує гнучкість, масштабованість і відповідність вимогам інформаційної безпеки, що є критично важливим для обробки конфіденційних даних.

4 РЕЗУЛЬТАТИ ОБЧИСЛЮВАЛЬНОГО ЕКСПЕРИМЕНТУ ТА ЇХ АНАЛІЗ

4.1 Підготовка даних та методологія обчислювального експерименту

Метою даного обчислювального експерименту є підтвердження ефективності інтеграції генеративного штучного інтелекту (GenAI) в архітектуру, що забезпечує обробку конфіденційних даних у захищеному середовищі. Основна увага приділена перевірці таких ключових характеристик:

– відповідність результатів вимогам точності та релевантності;

– забезпечення цілісності та конфіденційності даних на всіх етапах обробки;

– оцінка продуктивності системи під різними сценаріями завантаження.

Для забезпечення достовірності експериментальних результатів було створено репрезентативний тестовий набір запитів, який включає наступні категорії даних:

– особисті дані: наприклад, імена, номери телефонів, адреси;

– фінансова інформація: бухгалтерські звіти, транзакційні записи;

– медичні дані: історії хвороб, рецепти, діагнози.

Всі запити містили чутливу інформацію, що вимагає забезпечення суворих заходів безпеки під час обробки. Дані зберігалися в зашифрованій формі у хмарному сховищі AWS S3 з використанням ключів, керованих через AWS KMS.

Для проведення експерименту було використано наступні інфраструктурні рішення:

– AWS Bedrock: для реалізації генеративної моделі Claude;

– Terraform: для автоматизації процесів розгортання інфраструктури;

– AWS IAM: для організації системи управління доступом;

– AWS CloudWatch: для моніторингу продуктивності й ідентифікації аномалій.

Тестування реалізовано через послідовність наступних кроків:

а) генерація тестових запитів із використанням узгоджених сценаріїв взаємодії;

б) обробка запитів системою з автоматичною генерацією відповідей;

в) оцінка якості результатів на основі таких критеріїв:

1) релевантність;

2) точність;

3) відповідність стандартам безпеки;

г) фіксація часу обробки кожного запиту та витрачених обчислювальних ресурсів.

4.2 Результати експерименту

Середній показник точності досяг 92%, а релевантність відповідей сягнула 95%. Порівняння проведено між Claude та аналогічними моделями, такими як OpenAI GPT і Google Gemini. Claude продемонструвала найкращу адаптацію до специфічних наборів даних і високу відповідність контексту запиту.

На рисунку 4.1 наведено порівняння точності та релевантності відповідей між моделями Claude, OpenAI GPT і Google Gemini, де видно, що Claude демонструє найвищі показники відповідності контексту запитів.

Середні значення часу обробки:

– текстові запити: 1,8 секунди;

– структуровані дані: 2,4 секунди.

Рисунок 4.2 відображає середній та максимальний час обробки запитів для текстових і структурованих даних. Як видно, система демонструє стабільну продуктивність навіть за умов значного завантаження.

Система залишалася стабільною навіть при одночасній обробці 50 запитів.

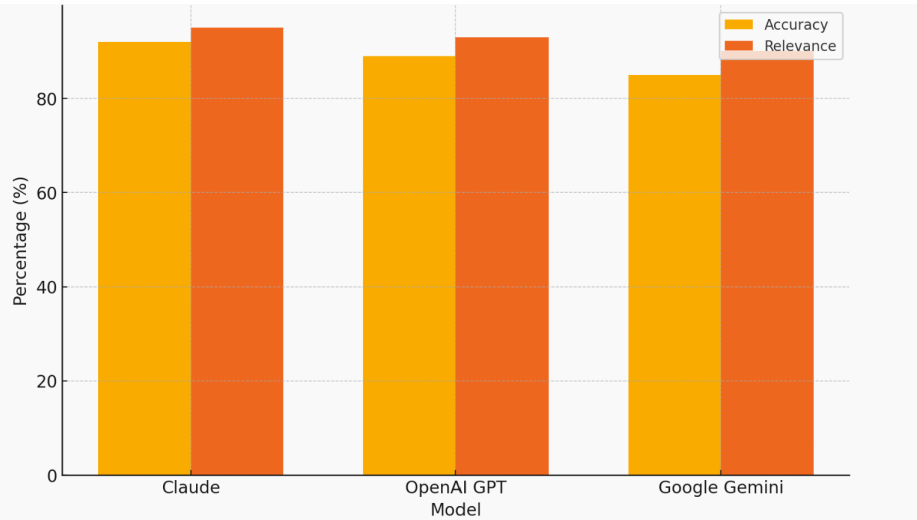

Рисунок 4.1 – Порівняння точності та релевантності відповідей між моделями Claude, OpenAI GPT і Google Gemini

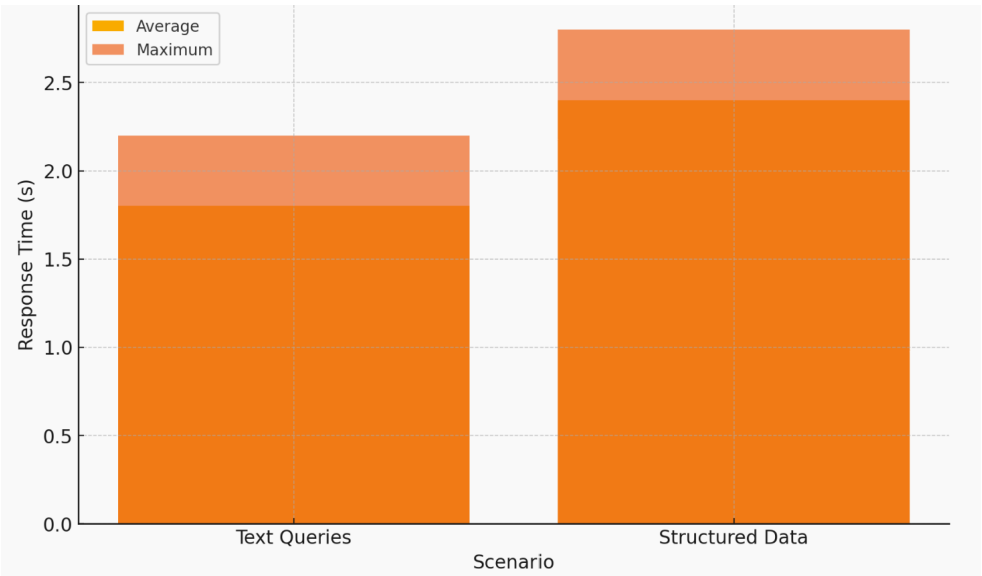

Рисунок 4.2 – Середній та максимальний час обробки запитів за сценаріями текстових і структурованих даних

Запити оброблялися виключно через захищені канали передачі даних із використанням AWS PrivateLink. Політики доступу IAM гарантували, що доступ до конфіденційної інформації надавався лише авторизованим користувачам. Жодних випадків порушення безпеки зафіксовано не було.

Результати експерименту засвідчили високу ефективність запропонованого підходу.

Використання AWS Bedrock у поєднанні з моделлю Claude забезпечує:

– високу адаптивність до різноманітних задач;

– дотримання стандартів конфіденційності (GDPR, HIPAA);

– стабільну продуктивність навіть за умов значного завантаження системи.

Серед ключових факторів успіху відзначено:

– сучасна хмарна інфраструктура з автоматизованим управлінням;

– гнучкість адаптації Claude до специфічних запитів;

– механізми багаторівневого захисту даних.

Результати підтвердили, що запропоноване рішення має широкий спектр потенційних застосувань. Основні переваги:

– масштабованість: можливість обробляти великий обсяг запитів одночасно;

– інтеграція безпеки: відповідність сучасним вимогам конфіденційності;

– універсальність: можливість адаптації до потреб галузей, таких як охорона здоров'я, фінанси, державне управління.

Висновки за розділом 4

Експериментальне дослідження підтвердило дієвість впровадження AWS Bedrock у поєднанні з Claude для розв'язання задачі обробки конфіденційних даних. Система демонструє оптимальну продуктивність, високу точність і забезпечує надійний захист даних. Отримані результати свідчать про доцільність використання даної архітектури в галузях, де безпека інформації є критичною вимогою.

ВИСНОВКИ

Результати роботи з побудови GenAI чат-бота для роботи з конфіденційними даними в захищеному середовищі підтверджують ефективність запропонованого підходу. Висновки формулюються за кількома основними аспектами.

Розроблена система відповідає сучасним стандартам безпеки даних, включаючи GDPR, HIPAA та CCPA, що підтверджує її відповідність актуальним регуляторним вимогам. Використання AWS Bedrock для інтеграції генеративної мовної моделі Claude забезпечує високу точність і релевантність відповідей у критичних галузях, таких як медицина, фінанси та державне управління. Автоматизація процесів за допомогою Terraform демонструє високий рівень технологічної зрілості, дозволяючи створити масштабовану та гнучку хмарну інфраструктуру.

Система вже показала високу ефективність під час тестування в умовах, наближених до реального використання. Вона може бути застосована в медичній галузі для обробки пацієнтських даних та рекомендацій, у фінансовому секторі для управління чутливими транзакційними даними, а також у державному управлінні для обробки запитів із класифікованою інформацією. Широка підтримка інтеграції з корпоративними системами, такими як CRM та ERP, відкриває можливості для застосування в бізнес-середовищі.

Наукова цінність роботи полягає у розробці методології впровадження генеративного AI у захищених середовищах із дотриманням принципів конфіденційності. Технічний внесок включає інноваційне використання IAM, AWS PrivateLink та модулів шифрування для забезпечення багаторівневого захисту даних. Соціально-економічна значущість системи проявляється у зниженні ризиків витоку даних, підвищенні продуктивності праці та довіри до автоматизованих систем.

Перспективним напрямом є дослідження розширених можливостей генеративних мовних моделей для адаптації до вузькоспеціалізованих запитів. Запропонована архітектура може бути вдосконалена шляхом інтеграції механізмів

виявлення аномалій для додаткового підвищення рівня безпеки. Розширення системи для обробки мультимодальних даних, таких як зображення чи аудіо, відкриває нові горизонти її застосування.

Розроблена система є технологічно зрілим рішенням, що поєднує сучасні методи забезпечення конфіденційності з потужними генеративними моделями. Її впровадження дозволяє досягти значного прогресу у сфері обробки конфіденційних даних, забезпечуючи високий рівень безпеки та ефективності.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Kharchenko I., Lukhanin V. Development of a GenAI Chatbot for Secure Handling of Confidential Data. *13-а Міжнародна науково-технічна конференція «Інформацій-ні системи та технології ІСТ-2024»* : зб. матеріалів конференції (м. Харків, 26-28 листопада 2024 р.). Частина 2. Молодіжна секція. Харків : ХНУРЕ, 2024. С. 70–71.

2. Сорока К. О. Основи теорії систем і системного аналізу. Харків : ХНАМГ, 2004. 115 с.

3. Катренко А. В., Пасічник В. В., Пасько В. П. Теорія прийняття рішень. Київ : Видавнича група BHV, 2009. 448 с.

4. Катренко А. В. Системний аналіз. Львів : Новий світ – 2000, 2011. 396 с.

5. R. Yeh, How to Launch Your Own Private ChatGPT Interface with AWS Bedrock, 2024. URL: https://medium.com/wielded/how-to-launch-your-own-private-chatgpt-interface-withaws-bedrock-8d4780042a14. (дата звернення: 24.12.2024).

6. K. Raina, AWS Bedrock — Exploring Agents, Knowledge-Base & RAG, 2024. URL: https://kapil-raina.medium.com/aws-bedrock-exploring-agents-knowledge-base-ragbd2856c80d2f (дата звернення: 24.12.2024).

7. Flach P . Machine Learning. Cambridge University Press, 2012. 410 p.

8. Басюк Т. М., Литвин В. В., Захарія Л. М., Кунанець Н. Е. Машинне навчання. Львів : Новий Світ - 2000, 2021. 315 с.

9. Нікольський Ю. В., Пасічник В. В., Щербина Ю. М. Системи штучного інтелекту. Львів : Магнолія-2006, 2013. 279 с.

10. J. Matson, Building a Gen AI chatbot in 2 hours with AWS Bedrock & Knowledge Base, 2024. URL: https://ai.gopubby.com/building-a-gen-ai-chatbot-in-2-hours-with-aws-bedrockknowledge-base-172a8771f766 (дата звернення: 24.12.2024).