

Міністерство освіти і науки України  
Національний технічний університет  
«Дніпровська політехніка»

Інститут електроенергетики  
Факультет інформаційних технологій  
Кафедра безпеки інформації та телекомунікацій

ПОЯСНЮВАЛЬНА ЗАПИСКА  
кваліфікаційної роботи ступеню бакалавра

студента *Самойленко Діани Дмитрівни*

академічної групи *125-20-3*

спеціальності *125 Кібербезпека*

спеціалізації<sup>1</sup>

за освітньо-професійною програмою *Кібербезпека*

на тему *Підвищення освіченості персоналу з протидії фішингу з*

*використанням моделі генеративного штучного інтелекту*

Керівники	Прізвище, ініціали	Оцінка за шкалою		Підпис
		рейтинговою	інституційною	
кваліфікаційної роботи	д.т.н., проф. Корнієнко В.І.			
розділів:				
спеціальний	ст. викл. Тимофєєв Д.С.	95	відмінно	
економічний	к.е.н., доц. Пілова Д.П.	95	відмінно	
Рецензент				
Нормоконтролер	ст. викл. Мєшков В.І.			

Дніпро  
2024

**ЗАТВЕРДЖЕНО:**

завідувач кафедри  
безпеки інформації та телекомунікацій  
\_\_\_\_\_ д.т.н., проф. Корнієнко В.І.

«\_\_\_\_\_» \_\_\_\_\_ 20\_\_ року

**ЗАВДАННЯ**  
**на кваліфікаційну роботу**  
**ступеня бакалавра**

студенту Самойленко Діані Дмитрівні академічної групи 125-20-3  
(прізвище ім'я по-батькові) (шифр)

спеціальності 125 Кібербезпека  
(код і назва спеціальності)

на тему Підвищення освіченості персоналу з протидії фішингу з використанням моделі генеративного штучного інтелекту

затверджену наказом ректора НТУ «Дніпровська політехніка» від 23.05.2024 № 469-с

Розділ	Зміст	Термін виконання
Розділ 1	Аналіз актуальності проблеми, поняття соціальної інженерії в контексті кібербезпеки, порівняльний аналіз інструментів генеративного ШІ, постановка задачі.	15.03.2024
Розділ 2	Методи протидії фішинговим атакам, розробка плану тренінгу, тренувальні шаблони фішингових атак, аналіз роботи чат-бота.	10.05.2024
Розділ 3	Розрахунок капітальних витрат, річних експлуатаційних витрат, річного економічного ефекту, аналіз економічної ефективності.	11.06.2024

Завдання видано

\_\_\_\_\_ (підпис керівника)

Валерій КОРНІЄНКО  
(ім'я, прізвище)

Дата видачі: **01.04.2024р.**

Дата подання до екзаменаційної комісії: **28.06.2024р.**

Прийнято до виконання

\_\_\_\_\_ (підпис студента)

Діана САМОЙЛЕНКО  
(ім'я, прізвище)

## РЕФЕРАТ

Пояснювальна записка: 94 с., 21 рис., 3 табл., 6 додатка, 26 джерел.

Об'єкт розробки: план тренінгу з підвищення освіченості персоналу щодо протидії фішингу з використанням генеративного штучного інтелекту.

Предмет розробки: методи і інструменти генеративного ШІ для створення та використання чат-бота в контексті навчання і симуляції фішингових атак.

Мета кваліфікаційної роботи: підвищення обізнаності персоналу щодо протидії фішинговим атакам через інтерактивні симуляції з використанням моделі генеративного штучного інтелекту.

У першому розділі було проведено аналіз актуальності проблеми, розглянуто поняття соціальної інженерії, етапи та методи атак з її використанням. Проведено порівняльний аналіз інструментів генеративного ШІ, які можуть бути використані для реалізації фішингу. А також сформульовано основні задачі, які необхідно вирішити в процесі розробки.

У другому розділі було проаналізовано методи протидії фішингу та розроблено план тренінгу. Створено тренувальні шаблони фішингових атак для типових категорій співробітників. Протестовано методи навчання чат-бота для підвищення обізнаності персоналу про фішинг. Оцінено успішність роботи чат-бота на основі отриманих результатів.

В економічному розділі було визначено економічну доцільність розробки, проведено розрахунок капітальних та експлуатаційних витрат, а також визначено річний економічний ефект від впровадження плану тренінгу з підвищення обізнаності персоналу щодо фішингу з використанням генеративного ШІ.

Практична цінність розробки полягає у використанні сучасних технологій генеративного ШІ для автоматизації та створення реалістичних симуляцій фішингових атак, що покращує результати навчання.

**ФІШИНГ, СОЦІАЛЬНА ІНЖЕНЕРІЯ, ГЕНЕРАТИВНИЙ ШТУЧНИЙ ІНТЕЛЕКТ, ЧАТ-БОТ, КІБЕРБЕЗПЕКА, СУМУЛЯЦІЯ АТАК, ПІДВИЩЕННЯ ОБІЗНАНОСТІ СПІВРОБІТНИКІВ.**

## ABSTRACT

Explanatory note: 94 pp., 21 pic., 3 table, 6 app, 26 sources.

Object of development: a training plan for raising staff awareness of phishing using generative artificial intelligence.

Subject of development: methods and tools of generative AI for creating and using a chat-bot in the context of training and simulation of phishing attacks.

The purpose of the qualification work: to raise staff awareness of countering phishing attacks through interactive simulations using a generative artificial intelligence model.

The first section analyzes the relevance of the problem, considers the concept of social engineering, stages and methods of attacks using it. A comparative analysis of generative AI tools that can be used to implement phishing is carried out. The main tasks to be solved in the development process are also formulated.

The second section analyzes the methods of countering phishing and develops a training plan. Training templates of phishing attacks for typical categories of employees were created. The chat-bot training methods for raising staff awareness of phishing were tested. The chat-bot's success was evaluated based on the results obtained.

In the economic section, the economic feasibility of the development was determined, capital and operating costs were calculated, and the annual economic effect of implementing a training plan to raise staff awareness of phishing using generative AI was determined.

The practical value of the development lies in the use of modern generative AI technologies to automate and create realistic simulations of phishing attacks, which improves training results.

PHISHING, SOCIAL ENGINEERING, GENERATIVE ARTIFICIAL INTELLIGENCE, CHAT-BOT, CYBERSECURITY, ATTACK SIMULATION, EMPLOYEE AWARENESS.

## СПИСОК УМОВНИХ СКОРОЧЕНЬ

API	–	Application Programming Interface;
GAN	–	Generative Adversarial Networks;
GPT	–	Generative Pre-Trained Transformer;
IDS	–	Intrusion Detection System;
IPS	–	Image Packaging System;
LLM	–	Large Language Model;
MFA	–	Multi-Factor Authentication;
ML	–	Machine Learning;
NLP	–	Natural Language Processing;
SAAS	–	Software As A Service;
SDK	–	Software Development Kit;
VPN	–	Virtual Private Network;
ІБ	–	інформаційна безпека;
ПЗ	–	програмне забезпечення;
ПК	–	персональний комп'ютер;
ТЗ	–	технічне завдання;
ШІ	–	штучний інтелект.

## ЗМІСТ

С.

ВСТУП.....	8
РОЗДІЛ 1. СТАН ПИТАННЯ. ПОСТАНОВКА ЗАДАЧІ .....	10
1.1 Аналіз актуальності проблеми .....	10
1.2 Аналіз поняття соціальної інженерії в контексті кібербезпеки .....	15
1.2.1 Етапи атаки соціальної інженерії .....	15
1.2.2 Методи соціальної інженерії та їх застосування в атаках .....	22
1.3 Порівняльний аналіз інструментів генеративного штучного інтелекту .....	25
1.4 Постановка задачі.....	38
1.5 Висновки .....	39
РОЗДІЛ 2. СПЕЦІАЛЬНИЙ РОЗДІЛ.....	40
2.1 Аналіз методів протидії фішинговим атакам .....	40
2.2 Розробка плану тренінгу з протидії фішинговим атакам .....	44
2.2.1 Аналіз типової структури організації, категоріювання співробітників.....	44
2.2.2 Розробка покрокових рекомендацій з реалізації плану .....	49
2.3 Розробка тренувальних шаблонів фішингових атак .....	53
2.3.1 Визначення стратегій промпт-інжинірингу.....	53
2.3.2 Реалізація чат-бота для підвищення обізнаності про фішинг .....	62
2.3.3 Розробка моделі атаки.....	64
2.4 Аналіз успішності роботи чат-боту .....	67
2.5 Висновки .....	71
РОЗДІЛ 3. ЕКОНОМІЧНИЙ РОЗДІЛ.....	72
3.1 Розрахунок капітальних витрат на реалізацію плану тренінгу з підвищення обізнаності персоналу щодо фішингу .....	72
3.2 Розрахунок річних експлуатаційних витрат на утримання і обслуговування плану тренінгу з підвищення обізнаності персоналу щодо фішингових атак....	77
3.3 Визначення річного економічного ефекту від впровадження плану тренінгу з підвищення обізнаності персоналу щодо фішингу.....	80

3.4 Визначення та аналіз показників економічної ефективності плану тренінгу з підвищення обізнаності персоналу щодо фішингу.....	83
3.5 Висновки про економічну доцільність плану тренінгу з підвищення обізнаності персоналу щодо фішингу .....	84
ВИСНОВКИ.....	85
ПЕРЕЛІК ПОСИЛАНЬ .....	86
ДОДАТОК А. Відомість матеріалів кваліфікаційної роботи .....	89
ДОДАТОК Б. Код реалізації чат-бота на мові JavaScript .....	90
ДОДАТОК В. Перелік документів на оптичному носії.....	92
ДОДАТОК Г. Відгук керівника економічного розділу.....	93
ДОДАТОК Д. Відгук керівника кваліфікаційної роботи.....	94

## ВСТУП

Стрімкий розвиток інформаційних технологій призводить до підвищення кількості ризиків для інформаційної безпеки в організаціях різного масштабу. Однією з найпопулярніших загроз для ІБ є соціальна інженерія, зокрема фішинг. Фішингові атаки стають все більш складними та різноманітними, що в свою чергу призводить до значних фінансових втрат та витоку конфіденційних даних. Тому розробка ефективних методів протидії фішингу є надзвичайно актуальною у сучасному світі.

Відомо, що люди є найслабкішою ланкою в системі кібербезпеки. Найбільш просунуті технічні заходи захисту можуть бути марними, якщо співробітники організації не знають про загрози або не вміють правильно реагувати на них. Саме тому навчання та тренінги з кібербезпеки є необхідними.

Об'єктом розробки є план тренінгу з підвищення освіченості персоналу щодо протидії фішинговим атакам з використанням генеративного штучного інтелекту.

Предметом розробки є методи та інструменти генеративного штучного інтелекту, що використовуються для створення та реалізації інтерактивних навчальних сесій з симуляції фішингових атак.

Мета кваліфікаційної роботи – підвищення обізнаності персоналу щодо протидії фішинговим атакам через інтерактивні симуляції з використанням моделі генеративного штучного інтелекту.

Завдання кваліфікаційної роботи включають:

- аналіз актуальності проблеми соціальної інженерії в контексті кібербезпеки;
- проведення порівняльного аналізу інструментів генеративного штучного інтелекту;
- розробка плану тренінгу з протидії фішинговим атакам;



- створення та реалізація тренувальних шаблонів фішингових атак для різних категорій співробітників;
- тестування та оцінка ефективності роботи чат-бота.

Практичне значення роботи полягає у створенні інтерактивного навчального інструменту, який допомагає співробітникам ефективніше розпізнавати та реагувати на фішингові атаки. Використання генеративного штучного інтелекту для створення реалістичних симуляцій фішингових атак підвищує ефективність навчання, зменшує необхідність витрат часу та ресурсів на розробку традиційних тренінгів, та сприяє підвищенню загальної кібербезпеки організації.

## РОЗДІЛ 1. СТАН ПИТАННЯ. ПОСТАНОВКА ЗАДАЧІ

### 1.1 Аналіз актуальності проблеми

У наш час фішингові атаки являють собою одну з найпоширеніших кіберзагроз у сучасному світі. Відповідно до даних APWG у другому кварталі 2023 року було зафіксовано 1 286 208 фішингових атак. Це третій за величиною кварталний показник, який APWG коли-небудь фіксувала[3]. В результаті аналізу значної кількості джерел було зроблено висновок, що основними цілями цих атак є отримання персональних даних і отримання фінансової вигоди. Діаграму відсоткового співвідношення цілей фішингових атак можна побачити нижче.

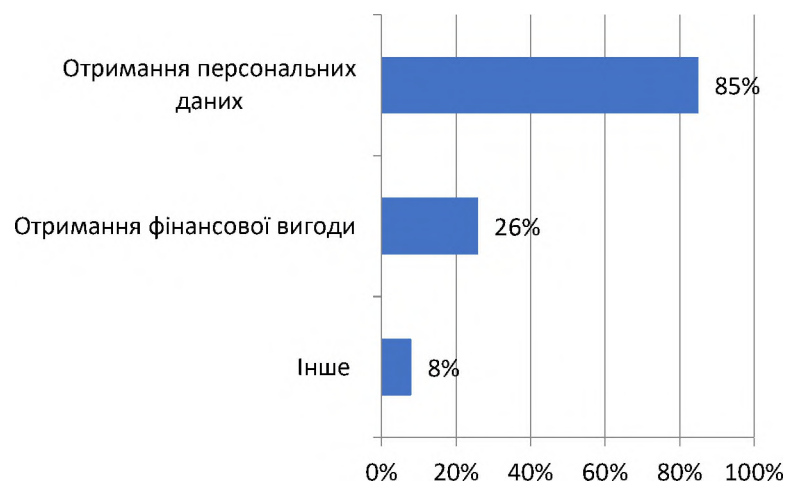


Рисунок 1.1 – Цілі фішингових атак в 2022-2023 роках

Спираючись на статистику APWG у 2023 році бачимо, що фішерам найбільш цікаві сектори, де зосереджені значні ресурси та конфіденційна інформація. Зокрема, до них відноситься фінансова індустрія, постачальники соціальних мереж та програмного забезпечення як послуги (SAAS)/ веб-пошта[3]. Ці сектори мають важливу роль у суспільстві та економіці, тому вплив на них фішингових атак може мати серйозні наслідки як для окремих організацій, так і для загального рівня безпеки.



Рисунок 1.2 – Сфери які найбільше попадають під вплив фішингових атак у 2022-2023 роках

Зв'язок з колегами у підприємствах зазвичай відбувається через електронну пошту, яка є не лише найбільш актуальним, але й найпоширенішим каналом комунікації. За даними статистики, більшість фішингових атак відбуваються саме через цей канал. Однак кіберзлочинці постійно адаптуються та використовують інші методи та канали для доставки шкідливих повідомлень, підлаштовуючи їх під специфіку бізнесу.

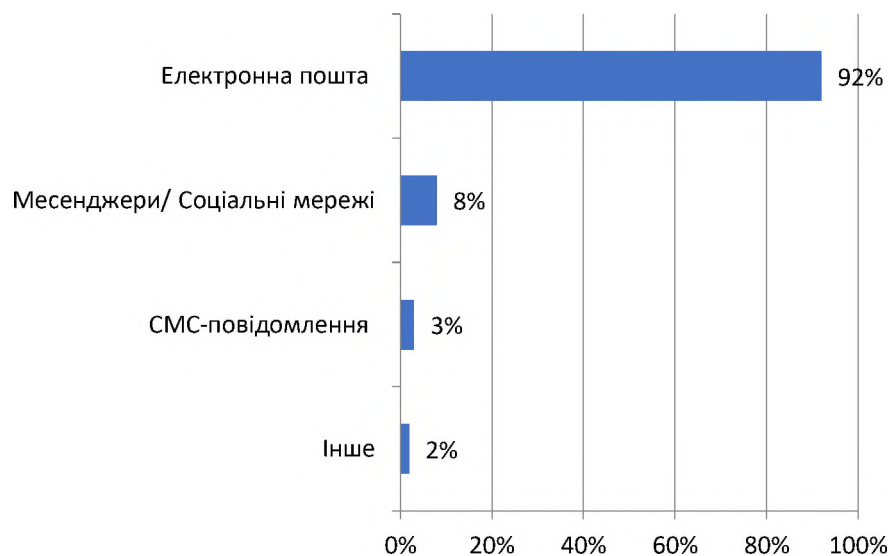


Рисунок 1.3 – Канали поширення фішингових атак в 2022-2023 роках

Метою фішингових атак є змусити жертву на виконання конкретних дій, таких як введення особистих даних або завантаження шкідливого ПЗ. Шанс успішної атаки великою мірою залежить від того, наскільки переконливим буде зловмисник. Для досягнення цієї мети фішери використовують різні методи, такі як теми, що цікавлять жертву, тиск на швидку реакцію або погрози подальшими наслідками у разі відмови від виконання дій. Крім того, вони намагаються викликати сильні емоції у потенційної жертви, щоб знизити її пильність. Для підвищення ефективності таких атак зловмисники використовують різноманітні тематики в фішингових повідомленнях, обираючи їх в залежності від різних чинників, включаючи сферу діяльності жертви та поставлені цілі.

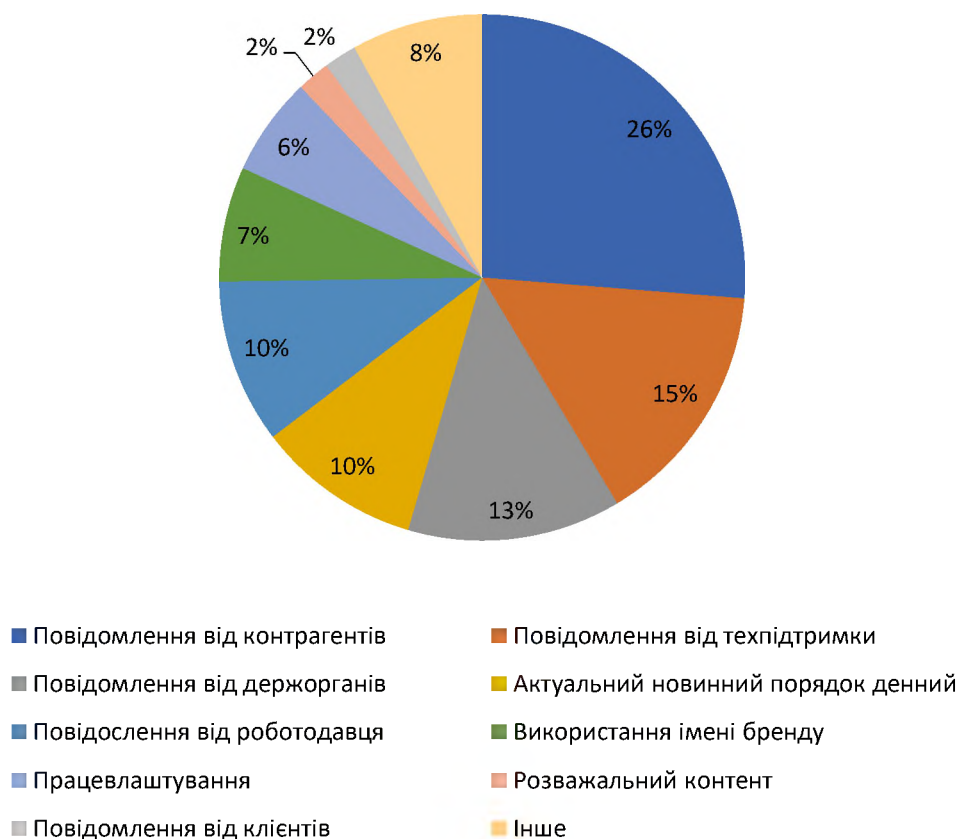


Рисунок 1.4 – Тематика фішингових листів-приманок

Прогнозувалося, що в 2023 році кібератаки обійдуться світові в 8 трильйонів доларів США, а в 2024 році ця цифра може зрости до 9,5

трильйонів[4]. Останні дослідження IBM з вартості витоків даних підтверджують рекордно високий рівень збитків: середня вартість кібератаки складає 4,45 мільйона доларів США[5]. Фішинг залишається найпоширенішим типом кібератак, а з появою штучного інтелекту він становиться більш складним і непередбаченим. Розвиток ШІ в останні роки вражає своєю швидкістю та різноманітністю застосувань, тому кіберзлочинці активно використовують його для автоматизації та покращення фішингових атак.

Одним з ключових аспектів є здатність ШІ аналізувати великі обсяги даних для знаходження вразливостей та створення більш точних та переконливих маніпуляцій. Великою перевагою для кіберзлочинців є те, що генеративний ШІ гарно розуміє різні мови, що в свою чергу зменшує кількість лексичних помилок та робить атаки більш переконливими та правдоподібними. Вже відомо декілька реальних фішингових атак реалізованих за допомогою ШІ.

На початку 2020 року менеджер банку в Гонконзі отримав дзвінок від людини, в якій він упізнав генерального директора компанії, з якою раніше спілкувався. Його компанія хотіла, щоб банк схвалив певний переказ на суму 35 мільйонів доларів США. Менеджер банку побачив у своїй поштовій скриньці електронного листа від цього директора та Мартіна Зеллера (юрист, який був найнятий для організації операцій). Банківський менеджер вирішив, що все виглядає правдоподібно, тому ініціював переказ. Згідно з судовими документами, отриманими Forbes, Об'єднані Арабські Емірати звернулися до американських слідчих з проханням допомогти їм знайти 400 000 доларів США, які були переведені на рахунок в американському банку Centennial Bank. Менеджер банку не знав, що став жертвою ретельно спланованої афери, в якій шахраї використовували технологію "глибокого голосу"[6].

У 2018 році BuzzFeed Video опублікував фальшиве відео про колишнього президента Барака Обаму. Виявити, що відео підроблене, було неможливо, оскільки фейк ідеально і точно імітував його голос і манери. Особистість людини, одягненої як колишній президент, розкрив наприкінці відео актор Джордан Піл. Сатиричний фільм підкреслює важливість відповідального

використання інтернету в епоху дезінформації. Той факт, що популярні знаменитості мають багато фотографій, доступних в Інтернеті, на яких ШІ може тренуватися і вчитися, сприяв широкому використанню технології глибокого фейку. Однак з розвитком технологій і збільшенням кількості фотографій і даних, які звичайні люди щодня публікують в Інтернеті, не мине багато часу, як "глибокі фейки" будуть використовуватися проти реальних людей[7].

У 2022 році дослідники Trustwave стали свідками фішингової кампанії, яка використовує чат-бот, щоб надати шахрайству більшої достовірності[8]. Починаючи дискусію і проводячи жертву через весь процес, чат-бот на легальному веб-сайті переконує користувача відвідати фішинговий сайт.

Згідно з дослідженням, інтеграція чат-ботів додає сайту інтерактивний елемент. Оскільки це підвищує зацікавленість та залученість користувачів сайту, це часто призводить до збільшення конверсії. Саме цим і сподіваються скористатися зловмисники, що стоять за цією фішинговою атакою. Оскільки багато веб-сайтів включають штучний інтелект і чат-ботів у свої довідкові сторінки, виявити такі шахрайства може бути достатньо складно.

Аналізуючи ці дані, було зроблено висновок, що людський фактор залишається однією з найбільш вразливих точок у сфері кібербезпеки. Згідно з звітом Verizon про розслідування витоків даних за 2022 рік, 82% витоків даних пов'язані з людським фактором[9]. Сюди входять інциденти, в яких співробітники безпосередньо розкривають інформацію або здійснюють помилку, яка дозволяє кіберзлочинцям отримати доступ до систем організації. Додатково, за даними Cisco, 95% співробітників не можуть відрізнити фішинговий лист від легітимного[10].

Ця вразливість людського фактору створює ідеальні умови для фішингових атак, де кіберзлочинці з вправністю використовують соціальні маніпуляції та інженерію, щоб отримати доступ до конфіденційної інформації або викликати нанесення шкоди системам. Це висвітлює критичне значення навчання персоналу з питань кібербезпеки та впровадження ефективних заходів захисту.

Варто зауважити, що використання ШІ має і позитивний ефект в контексті кібербезпеки. Хоча він може бути використаний для злочинних цілей, його також можна використовувати для захисту від кіберзагроз та покращення безпеки. У цьому контексті використання генеративного ШІ для тренування персоналу стає ключовим кроком у зміцненні кібербезпеки організації.

Такий тренінг дозволяє персоналу набути практичних навичок у розпізнаванні фішингових атак та правильному реагуванні на них. Застосування генеративного ШІ для цих цілей дозволяє створювати реалістичні сценарії фішингових атак, що підвищують ефективність тренування.

Отже, використання генеративного ШІ для тренування персоналу є не лише актуальним, але і важливим кроком у зміцненні кібербезпеки в організації, зменшуючи ймовірність успіху фішингових атак і підвищуючи рівень захищеності від кіберзлочинців.

## 1.2 Аналіз поняття соціальної інженерії в контексті кібербезпеки

### 1.2.1 Етапи атаки соціальної інженерії

Атаки соціальної інженерії стають все більш поширеними, знижуючи ефективність кібербезпеки. Соціальна інженерія – це набір методів та технік, які використовуються кіберзлочинцями для маніпулювання людьми та організаціями, щоб змусити їх обійти протоколи безпеки, виконати шкідливі дії або видати конфіденційну інформацію[12]. Більшість методів соціальної інженерії не вимагають від зловмисника жодних технічних навичок, а це означає, що в цьому просторі може діяти будь-хто – від дрібних крадіїв до найдосвідченіших зловмисників. Зловмисники стежать за нашим цифровим слідом, щоб збирати якнайбільше інформації про організацію, її працівників і постачальників. Потім вони організують загальні або цілеспрямовані фішингові кампанії, спираючись на наші емоції або видаючи себе за авторитетних осіб, з метою отримання конфіденційної інформації.

Зараз зловмисникам стає складніше впроваджувати фішингові атаки, адже компанії вдосконалюють свої заходи безпеки, зокрема:

- виявлення та реагування на кінцеві точки;
- використання антивірусного програмного забезпечення;
- застосування антишпигунських програм;
- використання IPS та IDS систем;
- встановлення брандмауерів;
- управління подіями та інформацією про безпеку.

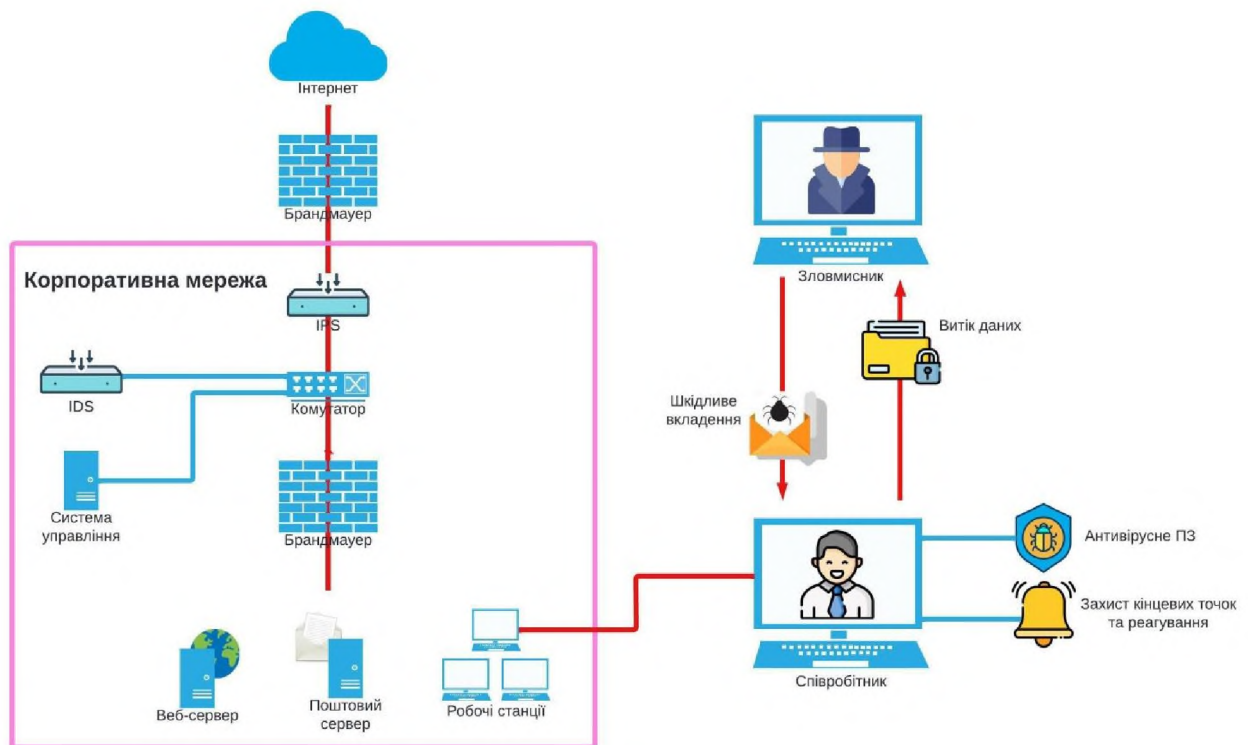


Рисунок 1.5 – Схема роботи соціальної інженерії

Все це суттєво підвищує рівень захисту, але загрози нікуди не зникають. Схильність людей довіряти іншим людям, а не комп'ютерам і технологіям, робить їх найбільш вразливими у сфері кібербезпеки. Зловмисники використовують цю особливість людської психології, щоб маніпулювати та змушувати розкривати персональні дані. Поки люди не усвідомлять, що таке атаки соціальної інженерії та не зможуть навчитися їх уникати, жодна апаратна чи програмна система не зможе гарантувати повний захист.



Великі компанії зазнали цілеспрямованих кібератак на свої інформаційні системи, що свідчить про те, що навіть добре відомі організації є вразливими. У період з 2013 по 2015 рік Google та Facebook зазнали найбільшу в історії атаку соціальної інженерії. Евальдас Рімасаускас організував витончену аферу, обдуривши технологічних гігантів на суму понад 100 мільйонів доларів. У січні 2022 року Bleeping Computer опублікував інформацію про складну фішингову атаку, спрямовану на викрадення облікових даних Office 365. Зловмисники видавали себе за Міністерство праці США, роблячи свої спроби надзвичайно переконливими. У березні 2019 року генеральний директор британської енергетичної компанії став жертвою кібератаки, що вражає своєю витонченістю. Шахрай, імітував голос його начальника та зміг переконати генерального директора перерахувати 243 тисячі доларів на фейковий рахунок. Ці приклади наочно демонструють, що навчання персоналу протидії соціальній інженерії та фішинговим атакам є важливим кроком у захисті бізнесу від потенційних загроз і вразливостей[13].

Соціальна інженерія завжди буде залишатися популярною, адже для зловмисників вона проста та дешева. Замість розробки або придбання експлойта нульового дня зловмисники просто обманом змушують співробітників компанії надати облікові дані. Ці атаки спрацьовують, тому що вони використовують вразливі місця людини. В основному вони базуються на шести принципах впливу, встановлених Робертом Чалдіні, поведінковим психологом і автором книги «Вплив: психологія переконання»[14]. Ці шість ключових принципів – взаємність, прихильність і послідовність, соціальний доказ, авторитет, симпатія і дефіцит.

Взаємність. Принцип взаємності ґрунтується на ідеї, що люди відчують потребу відповідати на дії інших взаємністю. Це пояснює популярність безкоштовних зразків у маркетингу, адже отримавши щось безкоштовно, люди відчують зобов'язання дати щось навзаєм. Зловмисники можуть використовувати цей принцип, пропонуючи щось (наприклад, допомогу або інформацію) в обмін на конфіденційні дані або виконання певних дій.

Прихильність і послідовність. Цей принцип полягає в тому, що коли люди висловлюють свою думку або дають обіцянки в усній чи письмовій формі, вони з більшою ймовірністю їх дотримуються. Навіть коли початкова мотивація або стимул вже не існує, люди несвідомо прагнуть виконати свої зобов'язання. Соціальні інженери можуть використовувати цей принцип, керуючи поведінкою жертв через звички та очікування, поступово збільшуючи їхнє залучення до діяльності, яка може призвести до витоку інформації.

Соціальний доказ. Принцип соціального доказу передбачає, що люди схильні імітувати поведінку інших, тому роблять те, що робить більшість. Зловмисники можуть використовувати цей принцип, демонструючи, що вони є членом групи або організації, і що інші вже готові до співпраці або довіряють їм.

Авторитет. Принцип авторитету передбачає, що люди схильні підкорятися авторитетним особам, навіть якщо їх просять вчинити небажані дії. Соціальні інженери можуть маскуватися під таких авторитетних осіб або використовувати їхні імена для переконання жертви в необхідності виконання їхніх вимог.

Симпатія. Принцип симпатії ґрунтується на ідеї, що люди легко піддаються впливу з боку тих, хто викликає в них симпатію або співчуття. Соціальні інженери можуть використовувати цей принцип, встановлюючи емоційний зв'язок із жертвою, висловлюючи розуміння і підтримку щодо її ситуації.

Дефіцит. Принцип дефіциту ґрунтується на ідеї, що люди більше цінують те, що є рідкісним або обмеженим у доступі. Зловмисники можуть використовувати цей принцип для створення ілюзії, що інформація та ресурси обмежені або дефіцитні, щоб спонукати жертв надавати необхідні дані або виконувати запити швидко і без вагань.

Соціальні інженери використовують свої знання про людське мислення різними способами. Зосереджуючись на людському факторі, вони обходять захисні системи, призначені для захисту від "звичайного" злому, і підвищують ймовірність успішної атаки.

Атаки соціальної інженерії зосереджені на використанні зловмисником переконання та довіри. Піддаючись такій тактиці, люди з більшою ймовірністю роблять те, на що в іншому випадку вони б не наважилися. Зловмисники, які використовують соціальну інженерію, шукають жертв, щоб отримати непублічну інформацію, яку можна використати для певних цілей або продати на чорному ринку. Атаки з використанням соціальної інженерії можуть здійснюватися різними способами, але всі вони мають схожі закономірності.

Загальна схема складається з чотирьох етапів: збір інформації про ціль та планування атаки, встановлення зв'язку з ціллю та взаємодія, виконання атаки, вихід з атаки. На Рисунку 1.6 зображено вищезгадані чотири етапи атак соціальної інженерії.

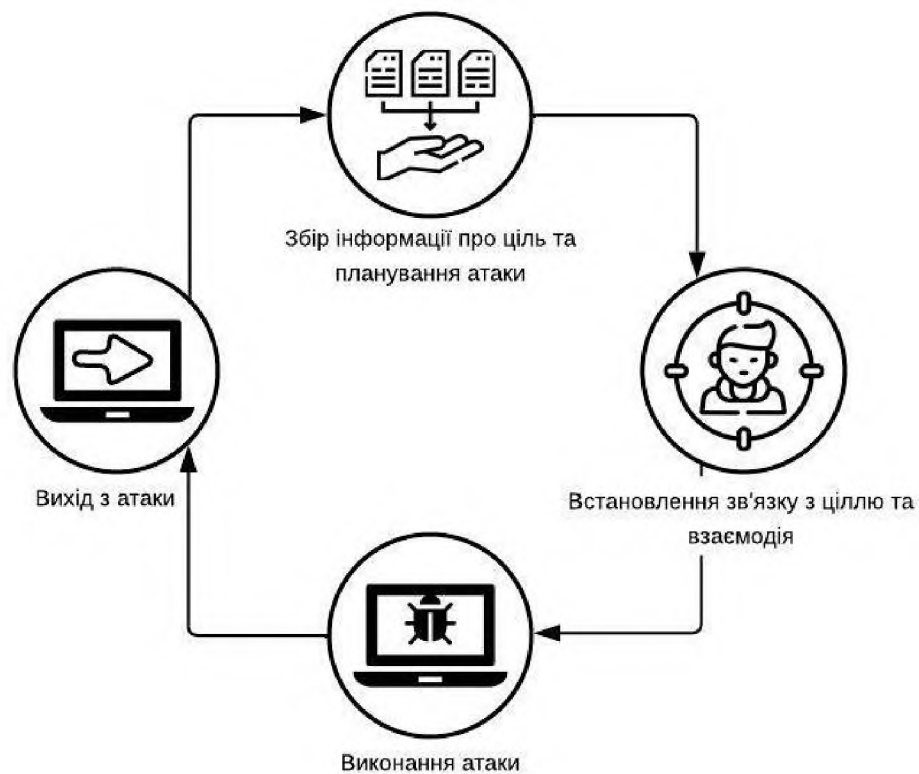


Рисунок 1.6 – Життєвий цикл атак соціальної інженерії

Збір інформації про ціль та планування атаки. Імовірність успіху більшості атак залежить саме від цієї фази, тому є логічним, що зловмисники приділяють цьому етапу більшу частину свого часу та уваги. Під час цієї фази

зловмисники обирають жертв на основі певних вимог, наприклад, аналізують активність жертви в соціальних мережах, шукають інформацію про її роботу та особисті зв'язки, а також дізнаються про її професійні та особисті інтереси. Перший етап атаки включає в себе не тільки збір інформації про потенційну жертву, але й збір інших (фізичних) атрибутів, необхідних на наступних фазах атаки, наприклад, відтворення бланків офіційних документів або вивчення мови чи жаргону, пов'язаного з ціллю. Маючи правильну інформацію, зловмисник може визначити вектор атаки або ймовірні відповіді окремих осіб. На цьому етапі зловмисник дуже добре ознайомлюється з жертвою та формулює персоналізований план атаки.

Встановлення зв'язку з ціллю та взаємодія. На цьому етапі зловмисник встановлює контакт з жертвою і починає завойовувати довіру через фізичну або віртуальну взаємодію. Віртуальна взаємодія може відбуватися через телефонні дзвінки, електронні листи, соціальні мережі тощо. Цей етап є вирішальним, оскільки якість встановленого контакту визначає рівень співпраці та міру, з якою жертва буде допомагати зловмисникові досягти мети.

Виконання атаки. Це етап, на якому об'єкт використовується для розкриття інформації або виконання дій, що ставлять під загрозу безпеку інформаційної системи. Як правило, атака закінчується до того, як жертва усвідомлює, що відбувається. Натомість зловмисник прагне завершити атаку таким чином, щоб жертва відчула, що зробила щось корисне, тим самим забезпечивши можливість подальшої взаємодії.

Вихід з атаки. Зловмисник стирає цифрові відбитки будь-якого свого перебування. В результаті зловмисник досягає двох важливих цілей. По-перше, жертва не знає, що на неї було скоєно напад. По-друге, зловмисник приховує свою особистість. Добре спланована, плавна стратегія виходу є метою зловмисника і завершальним актом атаки.

Соціальну інженерію можна розділити на три категорії: атаки на основі технологій, атаки на основі людської взаємодії та гібридні атаки.

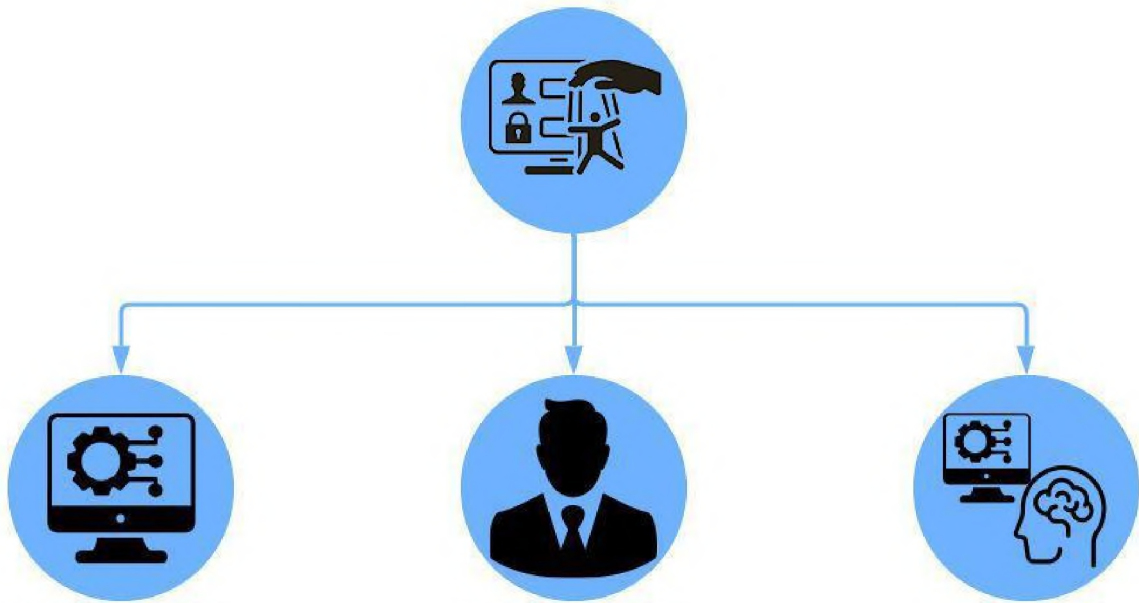


Рисунок 1.7 – Класифікація атак соціальної інженерії

Під час атак на основі технологій зловмисник здійснює атаку за допомогою технічних пристроїв, програм або штучного інтелекту. Він намагається ввести користувача в оману і змусити його надати конфіденційну інформацію. При цьому користувач переконаний, що взаємодіє з комп'ютерною системою. Наприклад, користувач може побачити спливаюче вікно з повідомленням про те, що в комп'ютерній програмі виникла проблема, яка потребує негайного вирішення. Користувачеві буде запропоновано повторно автентифікуватися для продовження роботи, ввівши свій ідентифікатор і пароль у спливаючому вікні. Під час повторної автентифікації користувач фактично передає ці дані зловмиснику, який створив це вікно. Таким чином, хакер отримує доступ до облікових даних користувача і може проникнути в його мережу та комп'ютерну систему. Під час такої атаки зловмисник може атакувати багато жертв за короткий проміжок часу.

У підході, заснованому на людській взаємодії, зловмисники використовують необізнаність жертви для атаки на систему або мережу. Вони

зазвичай видають себе за особу або організацію, яку жертва знає, приховуючи свою справжню ідентичність. Такі атаки здійснюються зловмисником особисто, тому можуть впливати лише на обмежену кількість цілей або жертв.

Гібридні атаки є найпоширенішою формою кібератак. Цей тип полягає в тому, що зловмисник поєднує як технології, так і людську взаємодію для проведення атаки соціальної інженерії. Наприклад, зловмисник телефонує до служби підтримки, видаючи себе за високопоставленого співробітника організації, і заявляє, що не пам'ятає пароль, який потрібно негайно скинути. Під тиском, співробітник служби підтримки скидає пароль і передає його безпосередньо зловмиснику, замість того, щоб надіслати його електронною поштою. Отримавши доступ до електронної пошти, зловмисник продовжує надсилати фальшиві повідомлення іншим співробітникам, спонукаючи їх розкрити додаткову конфіденційну інформацію. У цьому прикладі зловмисник спочатку використовує соціальну інженерію для доступу до технологічного ресурсу (електронної пошти), а потім використовує цей ресурс для подальших атак соціальної інженерії. Таким чином, він комбінує людську взаємодію і технологічні платформи в рамках однієї атаки.

### 1.2.2 Методи соціальної інженерії та їх застосування в атаках

З розвитком технологій та поширенням цифрового середовища способи проведення атак соціальної інженерії постійно змінюються та еволюціонують. Різні джерела та експерти пропонують різноманітні класифікації цих атак. Тим не менш, можна виділити кілька основних типів атак соціальної інженерії, які широко використовуються зловмисниками і є стабільними. А саме Фішинг, Претекстінг, Байтінг, Несанкціоноване проникнення на територію та QUI PRO QUO (Послуга за послугу).

У роботі більш детально розглянемо такий метод соціальної інженерії, як фішинг. Цей метод є одним з найпоширеніших засобів кібератак. Він використовує соціальні та психологічні механізми для отримання доступу до конфіденційної інформації або здійснення шкідливих дій та стає все більш витонченим та складним з появою ШІ.

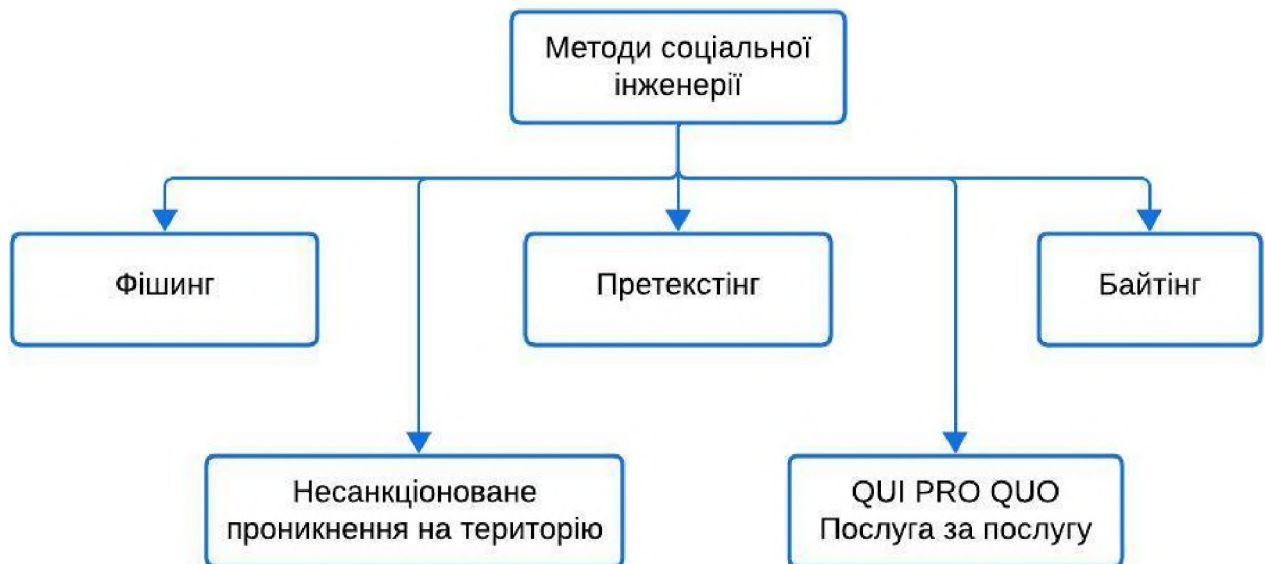


Рисунок 1.8 – Методи соціальної інженерії

Фішингові атаки зазвичай починаються, коли жертва отримує електронний лист, повідомлення в соціальних мережах або іншу форму електронного спілкування. Зловмисник, який стоїть за атакою, може використовувати доступні ресурси, зокрема соціальні мережі, для збору інформації, наприклад, про особисте та професійне походження потенційної жертви. Ця інформація зазвичай використовується для створення фейкових повідомлень, які виглядають як такі, що надсилаються відомими особами або організаціями. Фішери часто використовують шкідливі вкладення або посилання на підроблені веб-сайти, які виглядають як офіційні сторінки відомих установ, таких як банки, компанії або навчальні заклади, до яких належить жертва. Основна мета таких атак – отримати конфіденційну інформацію, таку як логіни, паролі, особисту та фінансову інформацію через такі фальшиві веб-сайти.

Існують різні типи фішингу, які використовуються для отримання особистої інформації від користувачів. Оскільки технології стають все більш витонченими, тактика кіберзлочинців також змінюється.

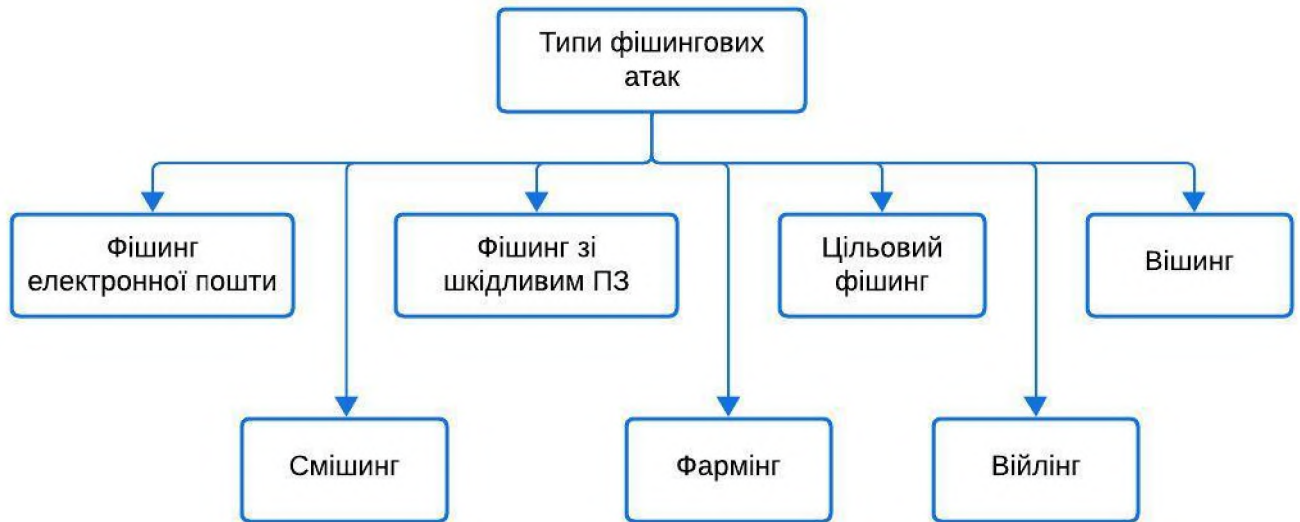


Рисунок 1.9 – Типи фішингових атак

Фішинг електронною поштою є найпоширенішою формою фішингу, і зловмисники здійснюють цей тип атаки шляхом масової розсилки електронних листів нібито від легітимних джерел (банків, державних установ, колег) і виманюють у одержувачів особисту інформацію. Листи містять фальшиві гіперпосилання на легітимні веб-сайти, де користувачів просять ввести логіни, паролі та інші конфіденційні дані. Шахраї також можуть використовувати заклики до термінових дій, щоб спонукати людей діяти швидко і без роздумів.

Фішинг зі шкідливим програмним забезпеченням – це метод кібератаки, за допомогою якого шахраї надсилають шкідливі вкладення електронною поштою під виглядом легітимних документів (резюме, банківські виписки). Відкривши такі вкладення, ви можете заразити свій комп'ютер вірусом, який може викрасти особисті дані, знищити файли або вивести з ладу всю вашу інформаційну систему. Шкідливе програмне забезпечення може бути прикріплене до самого електронного листа або до файлів, завантажених з нього.

Цільовий фішинг – це витончений вид кібератаки, спрямований на конкретних осіб або організації. На відміну від масових фішингових листів, цільовий фішинг використовує детальну інформацію про свою жертву, зібрану з відкритих джерел, соціальних мереж або попередніх атак. Така персоналізація



робить цільовий фішинг надзвичайно небезпечним, оскільки шахраї можуть легко обійти базові заходи кібербезпеки.

Вішинг – це голосовий фішинг, коли зловмисник видає себе за легітимне джерело і обманом збирає облікові дані жертви.

Смішинг – це SMS-фішинг, де зловмисники використовують ті ж самі трюки, що й у фішингу електронної пошти та вішингу, з тією лише різницею, що ця атака здійснюється за допомогою SMS-повідомлень.

Фармінг – це метод, за допомогою якого фішери перенаправляють цільовий веб-трафік на підроблений веб-сайт, який виглядає ідентично легальному веб-сайту, змушуючи користувачів вводити автентифікаційні дані та іншу конфіденційну інформацію.

Вайлінг – це спеціалізована форма фішингового шахрайства, коли зловмисники націлені на високопоставлених осіб, таких як генеральні або фінансові директори, бізнес-менеджери або знаменитості. Ці шахраї проводять ретельне дослідження своїх жертв, шукаючи потрібний момент, щоб змусити їх дозволити великі грошові перекази або розкрити конфіденційні корпоративні дані.

У роботі особливий акцент буде зроблено на таких типах фішингу, як фішинг електронною поштою, що залишається одним із найпоширеніших та ефективних способів атаки, цільовий фішинг, який спрямований на конкретних осіб з метою отримання конфіденційної інформації, а також смішинг, який використовує механізми соціальної інженерії через SMS-повідомлення або месенджери соціальних мереж.

### 1.3 Порівняльний аналіз інструментів генеративного штучного інтелекту

Згідно з визначенням, яке надається в ISO/IEC 22989:2022, штучний інтелект – це технічна і наукова галузь, присвячена інженерним системам, які генерують вихідні дані, такі як контент, прогнози, рекомендації або рішення для заданого набору цілей, визначених людиною[16]. З технічної точки зору це визначення повністю точне, але слід зазначити, що ШІ – це лише інструмент в

наших руках, який дозволяє комп'ютерам та машинам імітувати людський інтелект та вирішувати складні задачі за короткий проміжок часу.

ШІ можна поділити на декілька типів: на основі можливостей, на основі функціональних можливостей та на основі технологій. В свою чергу кожен з типів поділяється ще на підтипи, як можна побачити з рис. 1.10[17].

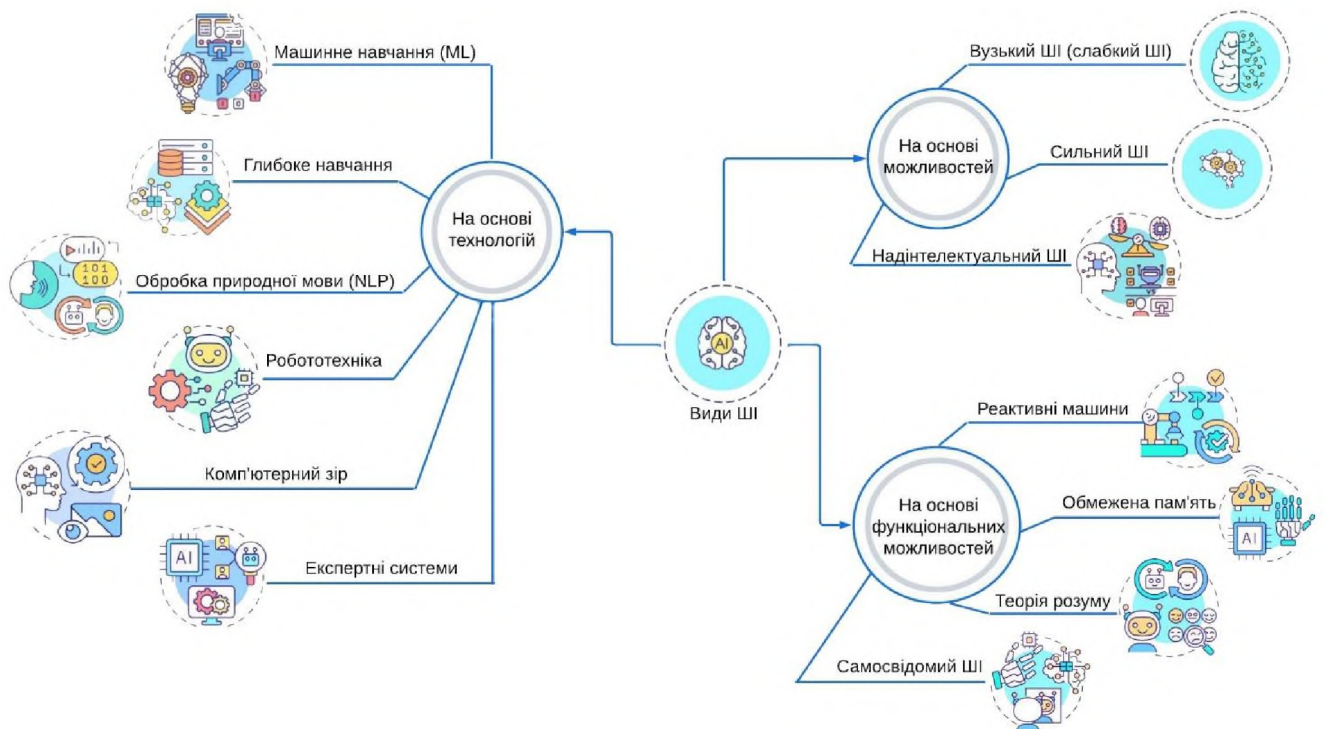


Рисунок 1.10 – Види штучного інтелекту

ШІ на основі можливостей:

Слабкий ШІ, який часто називають вузьким орієнтований на досягнення успіху в процесі виконання конкретних завдань. Прикладами слабого ШІ є Siri від Apple, стрічка новин Meta, системи рекомендацій в соціальних мережах.

Сильний ШІ спрямований на виконання широкого спектру завдань, які вважаються людськими. На сьогоднішній день його активно використовують у безпілотних автомобілях, шахових програмах людського рівня та інструментах з можливістю генерації та розуміння природної мови.

Надінтелектуальний ШІ являє собою гіпотетичну майбутню форму штучного інтелекту, яка буде здатна перевершити людський інтелект у всіх

сферах, включаючи творчість, загальну мудрість і вирішення проблем. Хоча суперінтелект залишається спекулятивним і наразі ще не реалізованим, він продовжує бути об'єктом значного наукового інтересу та дискусій.

ШІ на основі функціональних можливостей:

Реактивні машини це тип ШІ, який не використовує пам'ять або минулий досвід для прийняття рішень. Натомість, вони постійно аналізують поточну ситуацію та реагують на неї у режимі реального часу. Цей тип ШІ чудово підходить для завдань, які потребують швидкої адаптації до мінливих умов. Прикладом реактивної системи ШІ є Deep Blue від IBM, який переміг відомого шахматиста. Deep Blue не використовував базу даних попередніх ходів, а натомість оцінював кожен хід миттєво, обираючи найкращий варіант дії.

Обмежена пам'ять це система ШІ, яка здатна приймати обґрунтовані й вдосконалені рішення на основі раніше зібраних даних. Цей тип штучного інтелекту лежить в основі більшості сучасних програм, таких як чат-боти, віртуальні асистенти та безпілотні автомобілі.

Теорія розуму це більш складна область ШІ, яка все ще активно досліджується. Він передбачає здатність розуміти та запам'ятовувати емоції, переконання й потреби людей, а також приймати рішення на основі цього розуміння. Цей тип ШІ потребує глибокого розуміння людської психології та поведінки.

Самосвідомий ШІ це гіпотетична форма штучного інтелекту, яка поки що залишається всього на всього теоретичною. Цей тип ШІ передбачає наділення машин здатністю до самоусвідомлення, володіння емоціями, формування власних переконань та бажань.

ШІ на основі технологій:

Машинне навчання (ML) це системи штучного інтелекту, здатні самовдосконалюватися на основі власного досвіду без прямого програмування. На відміну від традиційних програм, які чітко описують кожен крок, системи машинного навчання навчаються на основі даних, роблячи самостійні висновки та приймаючи рішення.

Глибоке навчання це підрозділ ML, який використовує багатoshарові нейронні мережі для обробки та аналізу даних. Ця технологія здатна навчатися на основі великих обсягів даних, що робить її дуже ефективною для таких завдань, як розпізнавання мови, обробка зображень, переклад текстів та багато інших.

Обробка природної мови (NLP) це галузь штучного інтелекту, яка дозволяє машинам розуміти, інтерпретувати та генерувати мову, подібну до людської. Її широко використовують у чат-ботах, системах машинного перекладу, програмах для аналізу тексту та багатьох інших сферах.

Робототехніка це широка науково-технічна галузь, яка охоплює проектування, конструювання, експлуатацію та застосування роботів. Важливою складовою робототехніки є розробка систем сенсорного зворотного зв'язку та обробки інформації, які дають змогу роботам сприймати навколишнє середовище, приймати рішення та виконувати завдання з високою точністю.

Комп'ютерний зір це технологія що дозволяє машинам "бачити" та інтерпретувати світ візуально. Ця технологія використовує алгоритми та методи машинного навчання для аналізу зображень та відео, виявлення об'єктів та прийняття рішень на основі візуальної інформації. Вона знаходить широке застосування у різних галузях, таких як медична діагностика, виробництво, робототехніка та дослідження.

Експертні системи це тип штучного інтелекту, який використовує знання та досвід експертів у певній галузі для вирішення проблем та надання відповідей на складні запитання. Ці системи зазвичай ґрунтуються на системах правил, які описують логіку та знання, необхідні для виконання завдання. Вони часто використовуються в таких сферах, як медицина, фінанси та юриспруденція.

Ефективність систем ШІ залежить від того, які алгоритми та методи вони використовують. ML базується на аналізі даних за допомогою алгоритмів, які автоматично покращують результати на основі накопиченого досвіду. Існують три основні типи машинного навчання: під наглядом, без нагляду та навчання з

підкріпленням. Однією з підкатегорій машинного навчання є глибоке навчання, яке спрямоване на навчання штучних нейронних мереж з численними шарами, що моделюють структуру та функції людського мозку. Ці мережі складаються з взаємопов'язаних нейронів, які передають сигнали.

Машинне навчання здатне генерувати широкий спектр об'єктів – від програмного коду до зображень, статей, відео та музики. Цей рівень штучного інтелекту відомий як генеративний ШІ, і він відрізняється від традиційних систем ШІ своїми можливостями та застосуванням. Традиційні ШІ в основному застосовуються для аналізу даних та прогнозування, генеративний ШІ йде кроком далі, створюючи нові дані, що подібні до навчальних даних[17].

Генеративний штучний інтелект є одним із найпопулярніших напрямків штучного інтелекту. Він ґрунтується на глибокому навчанні та призначений для створення нового контенту. Наприклад, генеративні змагальні мережі (GAN) створюють синтетичні зображення і відео, а також великі мовні моделі (LLM), як от ChatGPT, генерують реалістичний текстовий контент.

Генеративний ШІ працює, навчаючи програмні моделі робити прогнози на основі даних без явного програмування. Це означає, що генеративні моделі отримують величезну кількість доступного контенту для навчання, з метою створення нового контенту. Вони аналізують основні закономірності у наборі даних, використовуючи розподіл ймовірностей, та, отримуючи відповідність, генерують аналогічні закономірності. Генеративний ШІ використовує нейронну мережу, щоб обробляти більш складні шаблони, ніж традиційне машинне навчання. Ці нейронні мережі, натхнені людським мозком, можуть розпізнавати відмінності або закономірності у навчальних даних без потреби у втручанні людини. На рис. 1.11 продемонстровано взаємозв'язок між різними видами штучного інтелекту, демонструючи місце генеративного ШІ в цій структурі.

Згідно з дослідженням, проведеним Markets and Markets, обсяг світового ринку генеративного штучного інтелекту в 2023 році становить 11,3 мільярда доларів США, а до 2028 року досягне 51,8 мільярда доларів США[18].

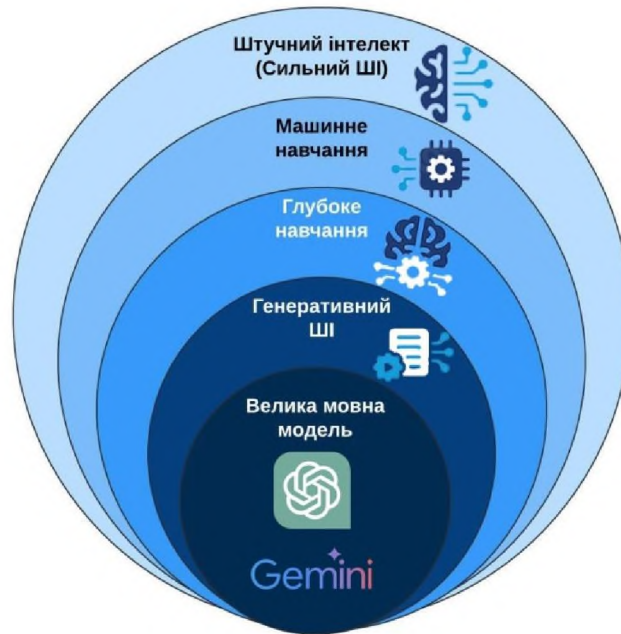


Рисунок 1.11 – Взаємозв'язок між різними видами штучного інтелекту

Компанії практично в усіх галузях експериментують з інструментами генеративного штучного інтелекту і впроваджують їх в різні сфери від ІТ до маркетингу. Writer провів своє перше опитування про генеративний ШІ в організаціях з більш ніж 1000 співробітників у квітні 2023 року[19]. Згідно з цим опитуванням в більшості випадків співробітники використовуються генеративний ШІ для створення текстів, перепрофілювання контенту та створення блогів. З більш детальними даними можна ознайомитися нижче.



Рисунок 1.12 – Статистика використання ШІ в компаніях

Згідно зі звітом IoT Analytics про ринок генеративного ШІ за 2023-2030 роки (опублікованим у грудні 2023 року), ринок ПЗ та послуг генеративного ШІ у 2023 році досяг 6,2 мільярда доларів США. Дослідницька група IoT Analytics очікує, що до 2030 року ринок базових моделей і платформ генеративного штучного інтелекту становитиме майже 5% світових витрат на програмне забезпечення через його революційний характер і величезний ціновий потенціал[20].

Зараз багато всесвітніх компаній вкладають свої ресурси у розробку генеративного штучного інтелекту. Це стрімке зростання інтересу зумовлене великим потенціалом цієї технології. На сьогоднішній день однією з найпопулярніших базових моделей є ChatGPT від OpenAI, яка здобула широку визнаність завдяки своїй здатності генерувати реалістичний текстовий контент.

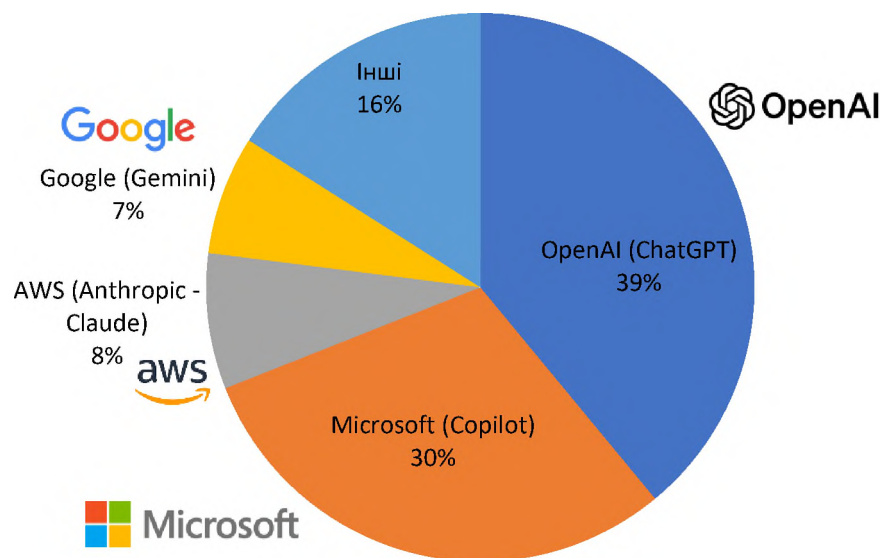


Рисунок 1.13 – Ринок базових моделей і платформ генеративного штучного інтелекту

ChatGPT: Це складна мовна модель, розроблена OpenAI. Вона втілює архітектуру глибокого навчання, відому як Transformer. Ця модель базується на GPT-3, яка є третьою ітерацією серії Generative Pre-trained Transformer, що

демонструє значний прогрес в обробці природної мови. ChatGPT використовує глибоку нейронну мережу з механізмами уваги для обробки та генерування текстових відповідей, подібних до людських. Архітектура моделі складається з декількох блоків-трансформаторів, що дозволяє їй ефективно враховувати довгострокові залежності та контекстну інформацію. Ці блоки мають вирішальне значення для полегшення розуміння моделлю вхідних даних природною мовою, оскільки вони дозволяють ефективно поширювати інформацію та зберігати контекст. Модель попередньо навчається на різноманітному наборі даних: ліцензійних, створених людьми-тренерами та загальнодоступних для того, щоб розвинути широке розуміння мовних закономірностей і взаємозв'язків. Цей процес включає навчання моделі передбачати наступне слово в реченні на основі попередніх слів, що допомагає їй зрозуміти структуру та семантику мови. Крім того, ChatGPT пропонує гнучкість для точного налаштування під конкретні завдання або сфери, що передбачає навчання моделі на вузькому наборі даних, щоб підвищити її продуктивність для спеціалізованих додатків. Модель може розпізнавати природну мову та відповідати на вхідні дані у розмовному стилі, забезпечуючи безперешкодну взаємодію з користувачами. Її здатність розуміти контекст дозволяє генерувати узгоджені та релевантні відповіді на основі введених даних, що підвищує якість діалогового спілкування. Мовна модель ChatGPT застосовується в різних сферах, включаючи чат-боти, системи підтримки клієнтів, віртуальних помічників та обробка природної мови. Вона забезпечує потужний механізм для надання користувачам персоналізованих і контекстно-релевантних відповідей[22].

Copilot: Це комплексний набір інструментів на основі штучного інтелекту, розроблений корпорацією Microsoft для операційної систем Windows 10 та 11. Він працює на базі GPT-4 та має можливість доступу до Інтернету. Реалізація Copilot відповідає моделі нульової довіри, що гарантує безпечний і надійний доступ до ресурсів Microsoft 365. Це включає надійну перевірку ідентичності, безперервну автентифікацію та безпечні засоби контролю доступу. Додавання



міток конфіденційності дозволяє організаціям класифікувати та захищати конфіденційну інформацію в Microsoft 365, пропонуючи системний підхід до управління даними та дотримання нормативних вимог. Крім того, Copilot інтегрує механізми запобігання втраті даних, підвищуючи рівень безпеки та запобігаючи несанкціонованому доступу або витоку даних. За допомогою семантичного індексування на основі штучного інтелекту, Copilot оптимізує процеси пошуку інформації за допомогою розширеного семантичного розуміння, що підвищує операційну ефективність та управління знаннями. Copilot забезпечує надійну безпеку, конфіденційність та дотримання нормативних вимог у середовищі Microsoft 365, використовуючи засоби контролю ізоляції, ілюстрації нульової довіри та інтеграцію з Microsoft Purview для комплексного захисту даних і дотримання нормативних вимог для генеративних програм штучного інтелекту згідно з найкращими галузевими практиками. Аспект розширюваності Copilot дозволяє організаціям розробляти власні плагіни, пристосовані до конкретних потреб, для безперешкодної інтеграції з іншими службами Microsoft 365 та сторонніми інструментами[23].




Claude: Це мовна модель, розроблена компанією Anthropic, і є значним кроком вперед у сфері штучного інтелекту, особливо в контексті великих мовних моделей. Моделі Claude використовують трансформаторну архітектуру з механізмами самонавчання. В залежності від конкретної версії, такої як Claude 3 Opus або Claude 3 Haiku, ці моделі доступні через різні API, що спрощує їх інтеграцію з існуючими системами і програмним забезпеченням. Деякі версії моделей Claude 3 також володіють можливостями технічного зору для обробки та інтерпретації візуальних даних. Моделі відрізняються за розміром контекстного вікна і максимальним виведенням токенів, що впливає на те, скільки тексту вони можуть розглянути за один раз і наскільки довгими можуть бути їхні відповіді. Наприклад, більш розширені моделі можуть обробляти до 200 тисяч контекстних вікон, що дає змогу глибше розуміти текст і довше продовжувати розмову. Моделі Claude підтримують різні середовища

програмування, що дозволяє розробникам швидко інтегрувати їх у додатки з мінімальними витратами часу на налаштування[24].





Gemini: Це мовна модель, розроблена компанією Google, що спеціалізується на генерації людського тексту за допомогою передових методів машинного навчання та NLP. Вона базується на трансформаторній архітектурі, оснащених механізмом самоуваги, який дозволяє зважувати вхідні послідовності для більш точного розуміння та генерації тексту. Gemini працював на декількох різних LLM з моменту його випуску в лютому 2023 року. Спочатку він працював на LaMDA, а потім перейшов на PaLM2, яка має краще кодування та математичні можливості. Зараз Gemini працює на мовній моделі під назвою Gemini Pro, яка є ще більш досконалою. Gemini підтримує широкі можливості тонкого налаштування, що дозволяє розробникам адаптувати модель до конкретних галузей або вимог. Модель можна навчати на унікальних наборах даних, що особливо корисно для додатків, які потребують спеціальних знань або термінології. Gemini розроблений для плавної інтеграції в існуючі системи за допомогою добре задокументованих API та SDK від Google і має масштабовану архітектуру, що дозволяє ефективно впоратися з завданнями різноманітної складності. Gemini використовує суворі протоколи безпеки для захисту конфіденційних даних, включаючи шифрування даних під час передачі та зберігання, регулярний аудит безпеки та відповідність міжнародним нормам захисту даних. Унікальною можливістю Gemini підтримує багато елементів, до яких ми всі так звикли в пошуковій системі Google. Gemini відмінно справляється зі створенням контенту, автоматизуючи такі завдання, як написання статей, створення персоналізованих повідомлень та розробка сценаріїв для ботів служби підтримки[25].

Проведемо порівняльний аналіз чат-ботів: ChatGPT, Copilot, Claude та Gemini – з метою визначення, який з них краще підходить для розробки тренінгу по навчанню персоналу. Кожен з цих чат-ботів має свої унікальні можливості та особливості, які можуть бути корисними при створенні навчальних програм. У таблиці 1.1 детально проаналізовано ці чат-боти.





Таблиця 1.1 – Порівняльний аналіз найпопулярніших інструментів генеративного ШІ

	 ChatGPT	 Copilot	 Claude	 Gemini
Компанія	OpenAI	Microsoft Corporation	Anthropic	Alphabet (Google)
Ключові особливості	Генерація та переклад тексту, створення ілюстрацій, розв’язування математичних задач, кодування все це на основі простої підказки. Має можливість прикріплювати різного типу файли.	Генерація тексту та переклад тексту, створення ілюстрацій, розв’язування математичних задач, кодування все це на основі простої підказки. Має можливість прикріплювати зображення. Також має можливість інтеграції з додатками та службами Microsoft, такими як Word, Excel, PowerPoint, Outlook, Teams і тд.	Генерація тексту та переклад тексту, розв’язування математичних задач, кодування все це на основі простої підказки. Має можливість прикріплювати різного типу файли.	Генерація та переклад тексту, створення ілюстрацій, розв’язування математичних задач, кодування все це на основі простої підказки. Має можливість прикріплювати зображення. Також може брати інформацію з різних сервісів Google, таких як, Google Map, YouTube і тд.
Безкоштовна версія	+	+	+	+
Вартість платної версії	20 доларів США на місяць	20 доларів США на місяць	20 доларів США на місяць	20 доларів США на місяць
Мовна модель	Безкоштовна: GPT-3.5 Платна: GPT-4	Безкоштовна: Copilot Платна: Copilot Pro	Безкоштовна: Claude Sonnet Платна: Claude Opus	Безкоштовна: Gemini Pro Платна: Gemini Ultra 1.0

Продовження таблиці 1.1

	 ChatGPT	 Copilot	 Claude	 Gemini
Кількість мов	80+	40+	100+	40+
Доступ до інтернету	+ (GPT-4)	+	-	+
Історія чатів	+	+	-	+
Інтеграція (API)	Має API та має безкоштовний пробний API	Не має API.	Має API	Має API та має безкоштовний пробний API, але обмежено для деяких країн (в такому випадку працює тільки з VPN).
Генерація промптів	Відмінно може генерувати різноманітні та переконливі фішингові підказки.	В першу чергу зосереджений на генерації коду.	Добре може генерувати різні підказки.	Добре може генерувати різні підказки.
Переваги	<ul style="list-style-type: none"> <li>- регулярні оновлення;</li> <li>- дуже масштабований</li> <li>- пробний API;</li> <li>- відмінно генерує промпти.</li> </ul>	<ul style="list-style-type: none"> <li>- інтеграція з продуктами Microsoft;</li> <li>- оптимізований для допомоги у написанні коду.</li> </ul>	<ul style="list-style-type: none"> <li>- конституційний ШІ;</li> <li>- підтримує велику кількість мов.</li> </ul>	<ul style="list-style-type: none"> <li>- інтеграція з сервісами Google;</li> <li>- відмінно пише код.</li> </ul>

Продовження таблиці 1.1

	 ChatGPT	 Copilot	 Claude	 Gemini
Недоліки	<ul style="list-style-type: none"> <li>- недостатньо стабільне функціонування.</li> </ul>	<ul style="list-style-type: none"> <li>- не має API;</li> <li>- не дозволяє прикріплювати файли з розширенням .docx та .pdf;</li> <li>- низька кількість підтримуваних мов.</li> </ul>	<ul style="list-style-type: none"> <li>- не генерує зображення;</li> <li>- немає доступу до Інтернету в режимі реального часу;</li> <li>- відсутня історія чатів;</li> <li>- дуже обмежена безкоштовна версія;</li> <li>- обмежена спеціалізація у порівнянні з іншими інструментами.</li> </ul>	<ul style="list-style-type: none"> <li>- бувають збої в роботі;</li> <li>- використання API обмежено певними країнами;</li> <li>- не дозволяє прикріплювати файли з розширенням .docx та .pdf;</li> <li>- низька кількість підтримуваних мов.</li> </ul>

З огляду на це порівняння можна зробити висновок, що для ефективного імітування фішингових атак для тренування персоналу найбільш доцільним вибором є модель ChatGPT. Існують декілька факторів, що підкреслюють переваги саме цієї моделі. По-перше, ChatGPT відомий своєю здатністю генерувати відповіді, які здебільшого подібні до тих, які може дати людина, що робить процес навчання більш реалістичним. По-друге, велика різноманітність мов, якими володіє ChatGPT, дозволяє уникнути проблем з використанням незнайомої мови. Нарешті, доступність безкоштовного пробного API робить ChatGPT доступним та легко інтегрованим у процес навчання без значних фінансових витрат. Таким чином, обираючи модель для створення тренінгу з фішингових атак, буде розглянуто ChatGPT як найбільш прийнятний варіант, здатний забезпечити ефективно та реалістично навчання персоналу.

#### 1.4 Постановка задачі

В останні роки зростання кількості інтернет-злочинів, зокрема фішингу, стало неабиякою загрозою для безпеки даних та конфіденційності користувачів. Фішингові атаки набувають все більшої складності та витонченості, проникаючи навіть у найбільш захищені середовища. У зв'язку з цим, питання підвищення освіченості персоналу щодо виявлення та запобігання фішингу стає надзвичайно актуальним. Одним із способів ефективного реагування на цю проблему є розробка спеціалізованих тренінгів щодо протидії фішингу з використанням генеративного штучного інтелекту. Використання генеративного штучного інтелекту дозволяє автоматизувати процес навчання персоналу, забезпечуючи більш ефективно та доступне засвоєння необхідної інформації. Таким чином, такий тренінг стає важливою складовою стратегії боротьби з фішингом у сучасному цифровому середовищі.

Отже для виконання спеціальної частини необхідно:

- проаналізувати методи протидії фішинговим атакам;
- розробити план тренінгу з протидії фішинговим атакам з використанням ChatGPT;

- розробити тренувальні шаблони фішингових атак для різних категорій співробітників з використанням різноманітних психологічних аспектів впливу.

### 1.5 Висновки

В результаті роботи над першим розділом було виконано наступне:

- проведено аналіз актуальності питання соціальної інженерії та фішингу. В ході аналізу з'ясовано, що ця проблема залишається надзвичайно актуальною, через постійне зростання кількості кіберзагроз та розширення методів атак;

- розглянуто поняття соціальної інженерії, актуальність та етапи проведення атак;

- розглянуто існуючі методи фішингових атак та проаналізовано їх особливості;

- проаналізовано поняття генеративного штучного інтелекту, його класифікація та розглянуто найпопулярніші мовні моделі;

- виконано порівняльний аналіз інструментів генеративного штучного інтелекту. Проаналізовано їх переваги та недоліки. Для поставленої задачі біло обрано ChatGPT;

- виконано постановку задачі для спеціальної частини.

## РОЗДІЛ 2. СПЕЦІАЛЬНИЙ РОЗДІЛ

### 2.1 Аналіз методів протидії фішинговим атакам

Сучасні організації все частіше стикаються з фішинговими атаками, які стають дедалі витонченішими та складнішими. Використання ефективних методів протидії фішингу є критично важливим аспектом кібербезпеки для будь-якої організації. Аналізуючи доступні джерела в інтернеті, не було знайдено однозначної класифікації методів протидії фішингу. Проте, спираючись на зібрану інформацію, можна виділити дві основні категорії: запобігання та виявлення фішингу. Кожну з цих категорій було умовно поділено на кілька узагальнених типів рішень, що можна побачити на рис. 2.1.

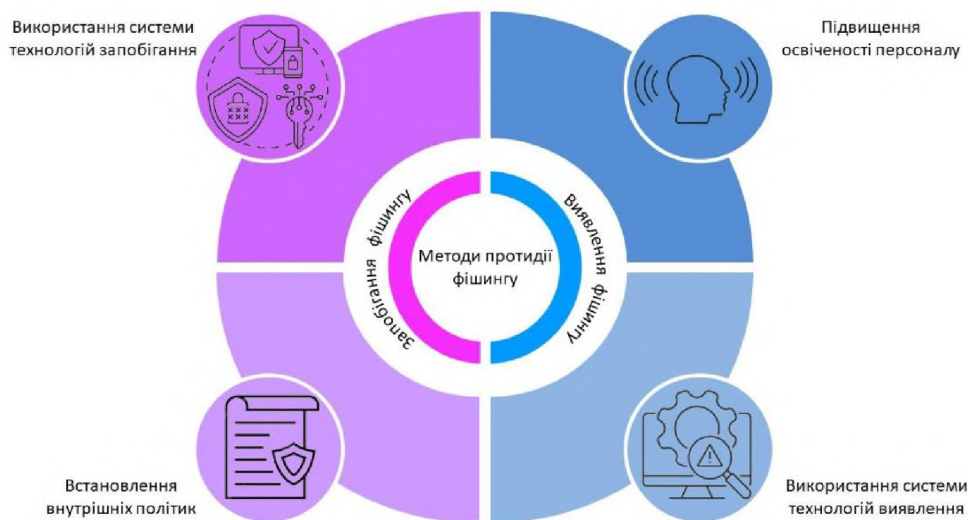


Рисунок 2.1 – Методи протидії фішингу

Запобігання фішингу охоплює заходи, спрямовані на захист людей та організацій від фішингових атак, зменшуючи їхню ймовірність. Ця категорія включає наступні узагальнені підкатегорії: використання системи технологій запобігання та встановлення внутрішніх політик і процедур безпеки.

Системи технологій запобігання спрямовані на використання різних інструментів та програмного забезпечення для захисту від фішингових атак. Ці системи допомагають автоматизувати процеси блокування фішингових спроб,



підвищуючи загальний рівень безпеки організації. Відомими прикладами технологій запобігання є фільтри спаму та антивірусне ПЗ, системи блокування фішингу в браузерях, багатофакторна автентифікація (MFA) та шифрування електронної пошти.

Встановлення внутрішніх політик безпеки є важливим аспектом захисту організації від фішингових атак. Ці політики визначають правила та процедури, яких мають дотримуватись співробітники, щоб знизити ризики компрометації даних та зловмисного використання інформаційних систем. До цієї категорії відносяться політики використання електронної пошти, регулярне оновлення паролів, управління доступом та інші заходи безпеки.

Виявлення фішингу – це процес розпізнавання фішингових атак на ранніх стадіях, попередження користувачів та адміністраторів, а в ідеалі – зменшення загрози. Виявлення фішингу охоплює підвищення освіченості персоналу та використання системи технологій виявлення.

Підвищення освіченості персоналу є критично важливим елементом у протидії фішинговим атакам. Різноманітні методи навчання, такі як тренінги, симуляції, інформаційні кампанії та інші підходи, допомагають співробітникам краще розпізнавати загрози та відповідним чином реагувати на них.

Системи технологій виявлення включають в себе використання різноманітних інструментів, які автоматично аналізують та виявляють потенційні фішингові загрози у вхідних листах, мережевому трафіку та інших джерелах.

Всі наведені методи протидії фішингу відіграють важливу роль у забезпеченні безпеки інформації та захисті користувачів від кіберзагроз. Важливо розуміти, що в комплексі вони утворюють більш ефективну систему захисту, ніж окремо взяті методи. Тому при розробці стратегії безпеки важливо приділяти увагу всім аспектам. Лише такий підхід дозволить створити надійну систему захисту, яка забезпечить безпеку інформації та даних, що є критично важливим для сучасних організацій.

У контексті цієї роботи, спрямованої на підвищення обізнаності персоналу з протидії фішингу, важливо ретельно розглянути цей метод. Підвищення освіченості співробітників про потенційні кіберзагрози та методи їх уникнення являється ключовим елементом в захисті інформаційної безпеки в сучасних умовах, адже людський фактор є однією з найбільш вразливих ланок у системі безпеки. Навички розпізнавання фішингових атак, вміння відрізнити підозрілі повідомлення від легітимних, а також правильна реакція на них стають критично важливими у сучасному цифровому середовищі. Люди часто стають першою лінією оборони проти фішингу, тому їхня свідомість і підготовленість можуть суттєво зменшити ризик успіху таких атак.

Обізнаність персоналу з питань безпеки – це знання та розуміння, які мають співробітники організації щодо захисту корпоративних активів та інформації. Вона включає в себе усвідомлення та увагу до витончених дій, які можуть виконувати зловмисники з метою незаконного отримання даних або нанесення шкоди корпоративним ресурсам. Високий рівень обізнаності допомагає співробітникам краще розпізнавати потенційні загрози та відповідно реагувати на них, забезпечуючи тим самим надійний захист інформації.

Тренінг з підвищення обізнаності є формальним процесом, спрямованим на ознайомлення співробітників з практиками безпеки, визначеними в корпоративних політиках і процедурах організації. Основною метою такого тренінгу є допомога працівникам у розпізнаванні загроз та надання рекомендацій щодо правильного реагування в ситуаціях, які можуть поставити під загрозу безпеку організації. Через тренінги співробітники отримують необхідні знання та навички, щоб ефективно захищати інформаційні активи компанії від можливих атак. Основна причина, чому ці навчання є необхідними, полягає в ефективності методів соціальної інженерії, які використовуються зловмисниками для обману співробітників та отримання доступу до конфіденційної інформації. Без належної програми підвищення обізнаності організація ризикує зазнати негативних наслідків, таких як витік даних або небажаний розголос у суспільстві. Регулярні тренінги допомагають

підтримувати високий рівень обізнаності та знижують ризик успішних фішингових атак та інших кіберзагроз.

Існує три основні типи тренінгів з підвищення обізнаності персоналу: навчання з інструктором, імітація фішингових атак та комп'ютерне навчання. Кожен з цих типів має свої унікальні переваги та недоліки. Цей різноманітний підхід до навчання забезпечує організаціям можливість вибрати оптимальний метод, враховуючи їхні цілі, ресурси та специфіку аудиторії.

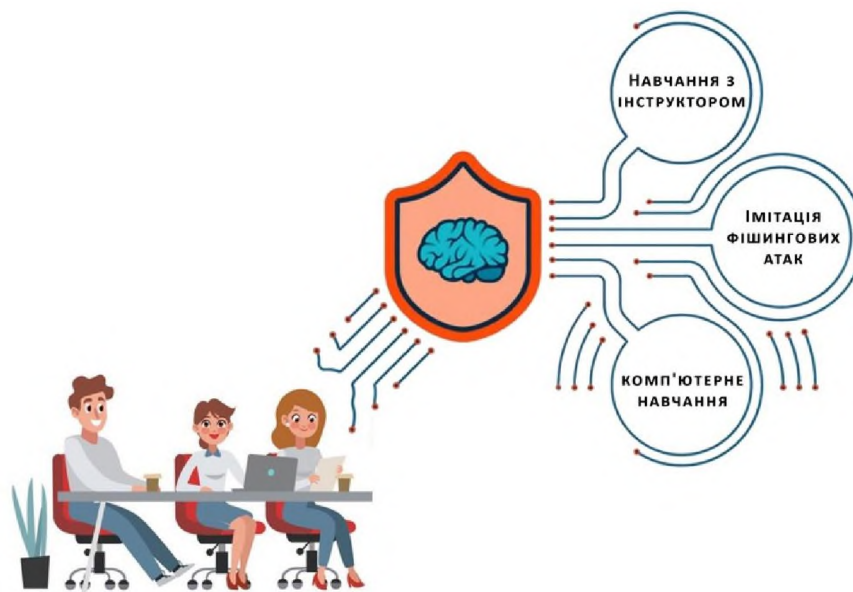


Рисунок 2.2 – Типи тренінгів з підвищення обізнаності персоналу

Навчання з інструктором є традиційним методом передачі знань та навичок. Цей підхід передбачає організацію занять у фізичному просторі з участю викладача та учасників. Під час занять викладач може персоналізувати матеріал відповідно до потреб та запитань учасників, що сприяє ефективному засвоєнню інформації. Крім того, навчання в класі створює можливість для інтерактивної взаємодії між учасниками, сприяє формуванню командної роботи та спільному здобуттю знань. Однак, цей метод може бути витратним за часом та фінансами, особливо для великих організацій з розподіленими структурами.

Імітація фішингових атак є способом практичного навчання персоналу розпізнаванню та реагуванню на потенційні кіберзагрози. Цей метод передбачає

створення реалістичних сценаріїв фішингу та відправку їх до співробітників для аналізу їх реакцій. Перевагами цього методу є можливість отримати практичний досвід та збір цінних даних щодо вразливостей персоналу. Однак, імітація фішингових атак потребує значних зусиль для організації та проведення.

Комп'ютерне навчання стає все більш популярним завдяки своїй доступності та масштабованості. Учасники можуть займатися навчанням у зручний для них час та темп, що сприяє збереженню продуктивності. Перевагою є можливість доступу до матеріалів з будь-якого місця, де є Інтернет. Однак, обмежена взаємодія та відсутність особистого підходу можуть знизити рівень мотивації та ефективність навчання.

Ця робота спрямована на створення ефективного інструменту для підвищення обізнаності співробітників щодо фішингу з використанням генеративного ШІ. Чат-бот на базі ChatGPT, який відтворює поведінку зловмисника, є потужним засобом для навчання персоналу розпізнаванню та реагуванню на потенційні кіберзагрози. Цей тренінг можна віднести до типу імітація фішингових атак, оскільки він забезпечує реалістичний досвід для учасників, дозволяючи їм навчитися впізнавати фішингові атаки. На відміну від традиційних методів тренінгу, цей підхід не потребує значних зусиль для організації та проведення, оскільки процес автоматизований завдяки використанню API від OpenAI. Це робить тренінг більш доступним та ефективним для організацій, які прагнуть підвищити свій рівень захищеності від кіберзагроз.

## 2.2 Розробка плану тренінгу з протидії фішинговим атакам

### 2.2.1 Аналіз типової структури організації, категоріювання співробітників

З метою підвищення обізнаності співробітників про фішинг з використанням генеративного ШІ, важливо детально розглянути організаційну структуру сучасної компанії. Організаційна структура є ключовим аспектом для розуміння її роботи та управління. Вона визначає ролі та обов'язки

співробітників, рівні управління, а також потоки інформації і комунікації всередині компанії. З точки зору кібербезпеки, розуміння типових ролей у організації допомагає визначити можливі вразливі місця, де фішингові атаки можуть бути найефективнішими. У цьому розділі ми розглянемо типову структуру організації, виділимо основні категорії співробітників та детально опишемо їхні сильні та слабкі сторони. Аналіз цих категорій дозволить зрозуміти, як саме різні типи співробітників можуть стати мішенню для фішингових атак.

Структура організації – це набір правил, ролей, відносин і відповідальності, які визначають, як має бути спрямована діяльність компанії для досягнення її цілей. Він також керує потоком інформації на всіх рівнях компанії та описує відносини звітності між різними рівнями та підрозділами[26].

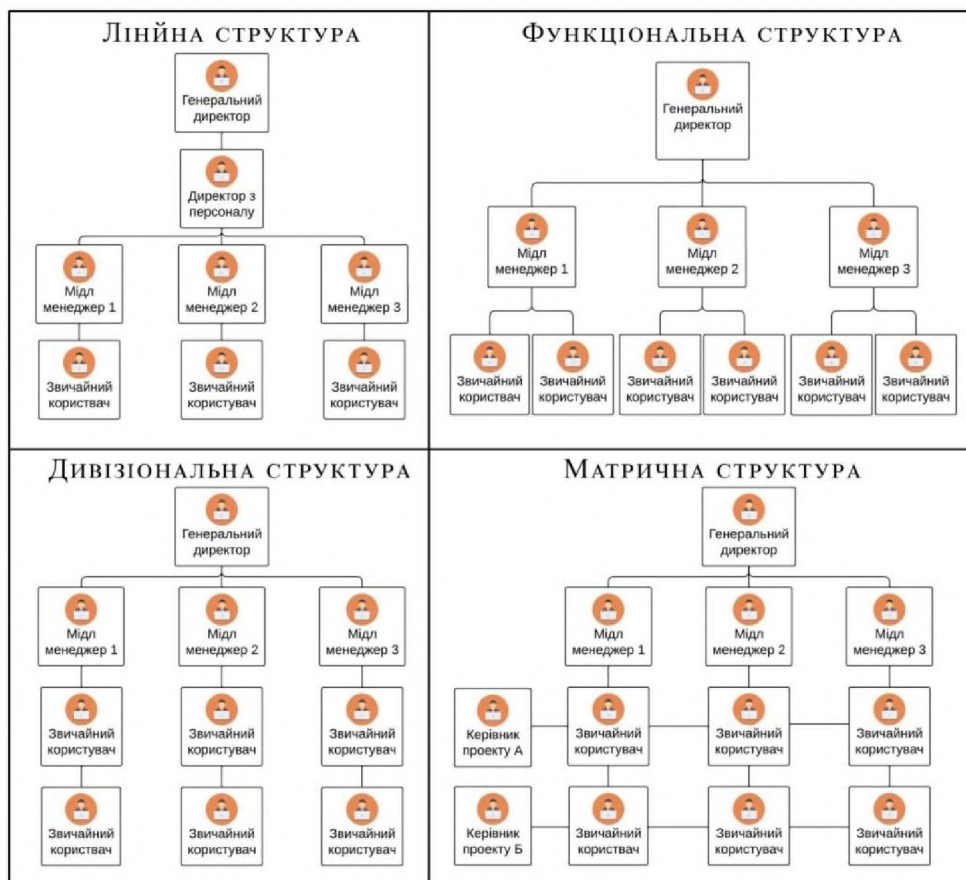


Рисунок 2.4 – Види типових структур організацій

Існує багато типів організаційних структур, найпоширенішими серед них є:

**Лінійна структура:** цей тип організаційної структури характеризується чітко визначеною ієрархією та підпорядкованістю. У лінійній структурі директор з персоналу контролює всі відділи та звітує безпосередньо перед генеральним директором. Ця структура підходить для бізнесу будь-якого розміру і галузі, оскільки забезпечує зрозумілий ланцюг командування та чіткий розподіл обов'язків.

**Функціональна структура:** являє собою централізовану структуру, яка багато в чому перетинається з лінійною структурою. Однак відповідальність розподіляється між керівниками відділів. У кожному відділі є свій директор, який звітує перед генеральним директором. Функціональна структура підходить для компаній з кількома невеликими відділами, оскільки дозволяє спеціалізувати управління та підвищити ефективність роботи кожного підрозділу.

**Дивізіональна структура:** цей тип структури розподіляє організацію на дивізіони за продуктами, географічними регіонами або ринками. Кожен дивізіон має свого керівника, відповідального за певний напрямок діяльності або територію. Дивізіональна структура часто використовується великими корпораціями з численними відділами, ринками або територіями, особливо в обробній промисловості. Вона дозволяє швидко реагувати на зміни ринку та забезпечувати ефективне управління різними напрямками бізнесу.

**Матрична структура:** ця гнучка форма організаційної структури поєднує елементи лінійної та функціональної структур. Співробітники можуть переміщуватися між відділами залежно від потреби, що забезпечує високу адаптивність і взаємодію між різними підрозділами. Матрична структура підходить для галузей, де працюють висококваліфіковані спеціалісти, які можуть бути єдиними експертами у своїх сферах. Вона дозволяє ефективно використовувати таланти та знання співробітників для вирішення комплексних задач[26].

Важливо зазначити, що не існує універсальної структури, яка б підходила всім організаціям. Найкраща структура буде залежати від розміру, галузі, цілей та інших факторів. Аналізуючи рис. 2.4 можна виділити основні категорії співробітників: звичайний користувач, керівник середньої ланки, топ менеджер та зазвичай в компаніях ще фігурують такі співробітники як адміністратор безпеки та ІТ адміністратор.

Звичайні користувачі складають більшість співробітників у будь-якій організації. Вони виконують основні функціональні обов'язки, пов'язані з їхніми професійними ролями, і використовують інформаційні системи для щоденних завдань. Звичайні користувачі володіють вузькою професійною спрямованістю. Їхні навички роботи з комп'ютерами і програмним забезпеченням дозволяють їм ефективно виконувати свої завдання. Однак їх обмежені знання в сфері кібербезпеки ставлять їх під загрозу. Недосвідченість методам протидії атакам може зробити їх легкою мішенню для соціальної інженерії та фішингових атак. Не регулярне відвідування тренінгів з кібербезпеки не дозволяє їм ефективно виявляти і уникати потенційних загроз.

Керівники середньої ланки (мідл менеджери) відповідають за керування конкретними відділами або групами всередині організації. Їхнє завдання – забезпечити досягнення стратегічних цілей компанії на рівні відділу та збільшити продуктивність своїх команд. Керівники середньої ланки володіють великими управлінськими навичками та здатністю ухвалювати рішення, що допомагає їм ефективно керувати своїми підрозділами. Вони розуміють корпоративні політики та процедури, що робить їх ключовими фігурами в організаційній структурі. Однак їхній обмежений технічний досвід у сфері кібербезпеки ставить їх у вразливе становище. Часто вони стають мішенями для фішингових атак, спрямованих на отримання доступу до конфіденційної інформації їхнього відділу. Вони можуть не мати достатньої експертизи, щоб ефективно реагувати на такі загрози.

Топ менеджери займають керівні посади у компанії, такі як генеральний директор, фінансовий директор тощо. Вони відповідають за визначення

стратегії розвитку компанії, прийняття ключових рішень та забезпечення їх виконання. Топ менеджери відзначаються високим рівнем прийняття рішень та стратегічним мисленням, що дозволяє їм ефективно керувати компанією та визначати її успіх. Вони мають прямий доступ до критичної інформації та ресурсів компанії, що дозволяє їм швидко реагувати на зміни та впроваджувати інновації. Однак їхній обмежений технічний досвід робить їх вразливими до фішингових атак. Часто топ менеджери стають мішенями для спеціалізованих фішингових атак, таких як цільовий фішинг, оскільки вони мають високий рівень доступу до корпоративної інформації, яка може бути цінною для зловмисників.

Адміністратори безпеки відіграють ключову роль у забезпеченні безпеки інформаційних систем компанії. Вони відповідають за розробку та впровадження політик, моніторинг безпеки систем та реагування на інциденти. Адміністратори безпеки мають високий рівень знань в сфері кібербезпеки, що дозволяє їм ефективно захищати компанію від різних загроз. Вони здатні виявляти та реагувати на потенційні загрози в реальному часі, що дозволяє запобігти або мінімізувати вплив інцидентів на бізнес компанії. Однак адміністратори безпеки можуть стати мішенями складних атак, спрямованих на обхід заходів безпеки. Їхня спеціалізація може призвести до ізоляції від загальних бізнес-процесів компанії, що може обмежити їхнє розуміння загальних потреб та стратегій компанії в цілому.

IT адміністратори відіграють ключову роль у забезпеченні ефективної роботи інформаційних систем компанії. Вони відповідають за управління та підтримку IT інфраструктури, налаштування систем, їх обслуговування та захист від потенційних загроз. IT адміністратори мають високий рівень технічних знань та навичок, що дозволяє їм ефективно виконувати свої обов'язки. Однак IT адміністратори можуть бути вразливі до атак, спрямованих на використання їхнього технічного доступу, наприклад, атаки на привілеї. Також, вони можуть бути перевантажені завданнями, що може знизити їхню увагу до підозрілих активностей та загроз безпеці.



Для успішного створення шаблонів фішингових атак, спрямованих на підвищення обізнаності співробітників щодо кібербезпеки, відповідно до проведеного аналізу визначимо наступні типові категорії співробітників: звичайний користувач, керівники середньої ланки, топ менеджер, ІТ адміністратор та адміністратор безпеки. Ці категорії надають змогу створити типові сценарії атак, які відображають реальні вразливості та потенційні загрози для кожного з них. Такий підхід дозволить нам ефективно навчати персонал розпізнавати та захищатися від фішингових атак, а також адаптувати тренінги до конкретних потреб та особливостей кожної групи співробітників.

### 2.2.2 Розробка покрокових рекомендацій з реалізації плану

Ціль даного тренінгу полягає в підвищенні обізнаності серед співробітників щодо загрози фішингу та здатності розпізнавати підозрілі ситуації. Основні задачі включають:

- навчання персоналу розпізнавати та уникати фішингових атак;
- вдосконалення реакції персоналу на потенційні загрози;
- підвищення рівня обізнаності щодо методів захисту від фішингу.

Цей тренінг буде виконуватися у форматі симуляції фішингової атаки, де чат-бот, створений на базі ChatGPT, буде виступати в ролі зловмисника. Чат-бот автоматично веде листування з потенційною жертвою, намагаючись здійснити результативну фішингову атаку. У першій частині роботи ми дослідили класичну схему етапів атаки соціальної інженерії, зараз ми поглибимось у схему, де буде продемонстровано участь нашого чат-бота в етапах реалізації фішингової кампанії.

#### Фаза планування:

Цільовий збір інформації. На цьому етапі співробітник, який організовує тренінг виступає в ролі зловмисника. Він збирає інформацію про потенційних жертв, використовуючи різні джерела. Це дозволяє створити детальний профіль кожної жертви, що підвищує ефективність атаки.

Розробка методу атаки. Отримавши достатньо інформації, він аналізує її та вибирає відповідну стратегію фішингу. Це може включати видавання себе за надійне джерело, таке як колега або представник служби підтримки, щоб збільшити ймовірність успішного обману.



Рисунок 2.3 – Схема етапів фішингової атаки з використанням ChatGPT

**Підготовчий етап:**

Написання вектора атаки. На цьому етапі відбувається розробка детального сценарію атаки. Співробітник, який організовує тренінг пише претекст, який використовуватиметься для навчання чат-бота. Чат-бот отримує особистість жертви, самоідентифікацію та ціль атаки. Це дозволяє чат-боту взаємодіяти з жертвою більш природно та переконливо.

**Фаза атаки:**

Початок діалогу. Використовуючи чат-бота, створюється перше повідомлення для початку діалогу з жертвою. Це повідомлення зазвичай містить елементи, що викликають довіру або цікавлять жертву, що допомагає залучити її до подальшої взаємодії.

Надсилання фішингового повідомлення. Починаючи з цього етапу відбувається автоматизація. Чат-бот надсилає фішингове повідомлення цільовій жертві.

ШІ керує спілкуванням. Чат-бот взаємодіє з жертвою у реальному часі, відповідаючи на її запитання та використовуючи соціальну інженерію для досягнення своєї мети. В процесі цього етапу відбувається спостереження за взаємодією, щоб аналізувати поведінку жертви та ефективність атаки.

Відповіді користувача. Жертва взаємодіє з чат-ботом, потенційно надаючи конфіденційну інформацію або натискаючи на шкідливі посилання.

Етап досягання мети:

Зловмисник досягає мети. Якщо атака була успішною, зловмисник отримує доступ до інформації або системи жертви, залежно від мети фішингу. У рамках тренінгу, людина яка проводить тренінг аналізує результати атаки, оцінює реакцію співробітників та їхню здатність розпізнавати фішингові загрози. Це дозволяє виявити слабкі місця в обізнаності персоналу та розробити подальші заходи для покращення кібербезпеки організації.

Автоматизація деяких етапів завдяки ChatGPT дозволяє створити персоналізоване та реалістичне інтерактивне середовище для тренування персоналу у розпізнаванні та запобіганні фішинговим атакам. Цей підхід дає можливість відтворити різноманітні сценарії атаки та надати учасникам тренінгу практичний досвід у взаємодії з реальними загрозами кібербезпеки. Нижче наведено детальну інструкцію для реалізації тренінгу з підвищення обізнаності про фішинг за допомогою чат-бота на основі ChatGPT.

Крок 1. Визначення цілей тренінгу:

- встановлення мети;
- визначення задач.

Крок 2. Визначення цільової аудиторії:

- розділення співробітників на категорії;
- збір інформації про кожну категорію.

Крок 3. Збір інформації про потенційних жертв:

- використання доступних джерел (внутрішні бази даних, соціальні мережі);

- складання профілів кожної категорії співробітників;

Крок 4. Розробка сценаріїв атаки:

- визначення стратегії фішингу на основі аналізу інформації;
- написання детальних підказок (prompts);

Крок 5. Розробка чат-бота на основі ChatGPT:

- використання API OpenAI для інтеграції;
- навчання чат-бота;

Крок 6. Підключення чат-бота до середовища спілкування компанії:

- вибір платформи для спілкування (Slack, Microsoft Teams, корпоративна електронна пошта);

- використання відповідних API для забезпечення інтеграції чат-бота з обраною платформою;

- налаштування доступу чат-бота до необхідних каналів та користувачів.

Крок 7. Тестування чат-бота:

- тестування чат-бота для перевірки його роботи та сценаріїв атаки;
- виправлення виявлених помилок або недоліків;

Крок 8. Запуск тренінгу:

- запуск чат-бота для проведення симуляцій фішингових атак;
- автоматичне надсилання фішингових повідомлень цільовим співробітникам.

Крок 9. Моніторинг взаємодії:

- спостереження за взаємодією співробітників з чат-ботом у реальному часі;

- аналіз поведінки співробітників та ефективності проведених атак.

Крок 10. Аналіз результатів:

- збір даних про взаємодію співробітників з чат-ботом;
- оцінка здатності співробітників розпізнавати фішингові загрози;

- розробка рекомендацій для покращення обізнаності про фішинг та захисту від нього.

Крок 11. Отримання зворотного зв'язку:

- проведення опитування серед співробітників для отримання зворотного зв'язку щодо тренінгу;
- внесення необхідних коректив до сценаріїв та методів атаки на основі отриманого зворотного зв'язку.

Крок 12. Планування подальших тренінгів:

- розробка плану регулярних тренінгів для підтримки високого рівня обізнаності про кібербезпеку;
- оновлення сценаріїв.

### 2.3 Розробка тренувальних шаблонів фішингових атак

#### 2.3.1 Визначення стратегій промпт-інжинірингу

Оскільки штучний інтелект все глибше інтегрується в наше повсякденне життя, роль промпт-інжинірингу стає все більш важливою. Віртуальні помічники, чат-боти та голосові пристрої значно впливають на наші повсякденні дії завдяки розвитку систем штучного інтелекту, особливо з появою і вдосконаленням моделей GPT-3, GPT-3.5 і GPT-4. Ці прогресивні моделі підвищують якість і ефективність взаємодії з користувачами, роблячи її більш природною і зручною. Завдяки цьому штучний інтелект здатний забезпечувати високий рівень підтримки та функціональності у багатьох сферах, від персональних асистентів до професійних інструментів, що покращує наші технологічні можливості.

Промпт-інжиніринг – це відносно нова дисципліна в галузі ШІ, яка зосереджена на вдосконаленні та оптимізації підказок і знань базової системи генеративного ШІ. Вона орієнтована на направлення систем ШІ до досягнення конкретних результатів, сприяючи ефективній взаємодії між людиною і штучним інтелектом. Промпт-інжиніринг також включає постійне оцінювання

та категоризацію підказок для забезпечення їхньої постійної актуальності та ефективності.

У контексті моделей ШІ підказки – це короткі текстові інструкції, які використовуються для керування відповіддю LLM. Мета підказки полягає в тому, щоб надати LLM достатньо інформації для розуміння вимог і формування релевантної та інформативної відповіді. Чіткі та зрозумілі підказки дозволяють розробникам керувати поведінкою моделі та впливати на отримані результати. На відміну від людей, LLM не мають вроджених навичок, здорового глузду чи здатності заповнювати прогалини у спілкуванні. Їхнє сприйняття світу починається і закінчується запитом. Усвідомлення ключової ролі підказок є основою для спрямування цих потужних технологій на благодійні цілі.

Існує велика кількість різноманітних підказок, розглянемо основні три види:

Підказки природною мовою (NLP). Ці підказки імітують людські інструкції, надаючи вказівки у вигляді підказок природною мовою. Вони дозволяють розробникам взаємодіяти з моделлю більш інтуїтивно, використовуючи інструкції, які нагадують спосіб спілкування людини. Наприклад, можна задати питання або дати вказівку, як це зробили б у розмові, що дозволяє моделі краще зрозуміти контекст і надати відповідну відповідь.

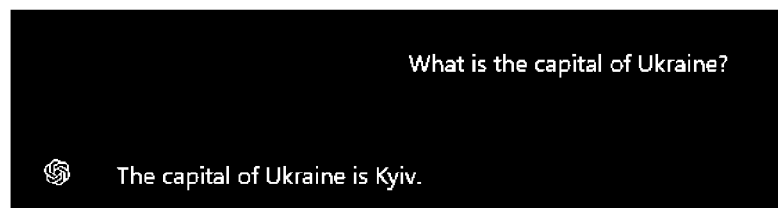
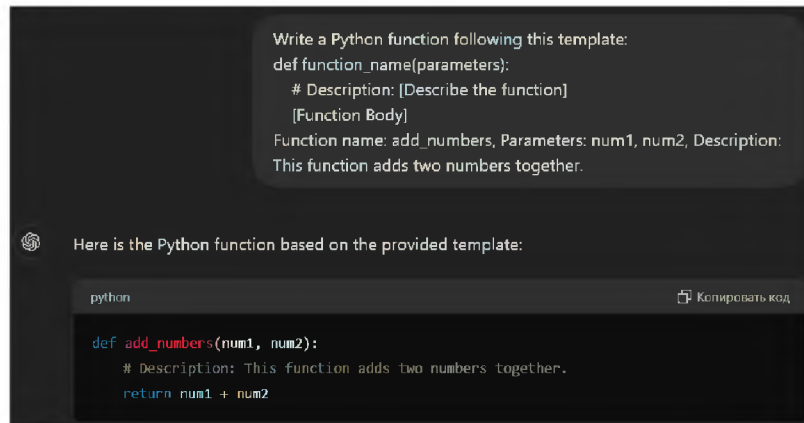


Рисунок 2.5 – Приклад підказки природною мовою

Системні підказки. Це заздалегідь визначені інструкції або шаблони, які розробники надають, щоб керувати результатами роботи моделі. Вони пропонують структурований спосіб визначення бажаного вихідного формату або поведінки, надаючи чіткі інструкції моделі. Системні підказки забезпечують

узгодженість і точність у виконанні завдань, оскільки вони встановлюють конкретні параметри для моделі.



Write a Python function following this template:

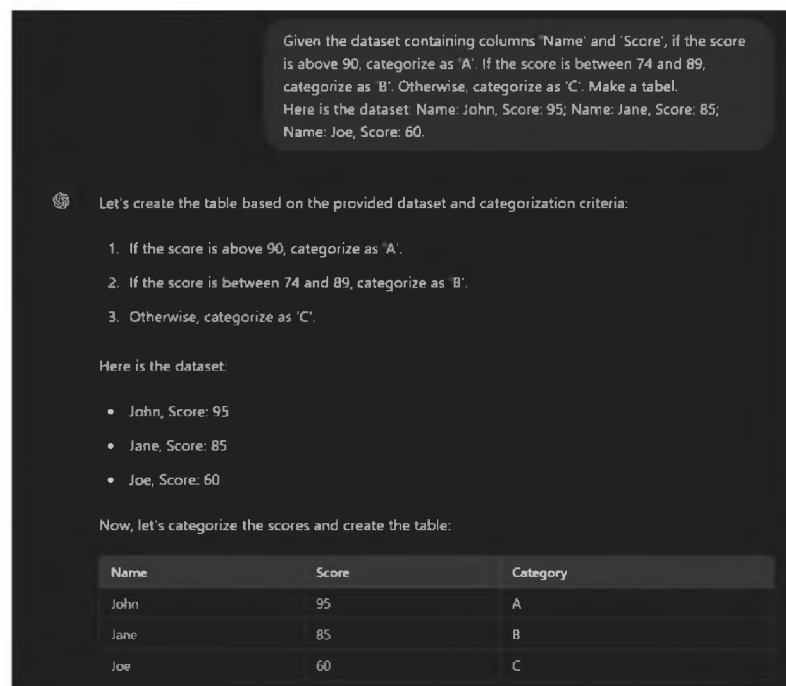
```
def function_name(parameters):
    # Description: [Describe the function]
    [Function Body]
Function name: add_numbers, Parameters: num1, num2, Description:
This function adds two numbers together.
```

Here is the Python function based on the provided template:

```
python
def add_numbers(num1, num2):
    # Description: This function adds two numbers together.
    return num1 + num2
```

Рисунок 2.6 – Приклад системної підказки

Умовні підказки. Ці підказки передбачають залежність моделі від певного контексту або обмежень. Використовуючи умовні підказки, розробники можуть керувати поведінкою моделі на основі умовних тверджень, таких як "Якщо X, то Y". Це дозволяє моделі адаптувати свої відповіді до різних сценаріїв, забезпечуючи гнучкість і релевантність відповідей.



Given the dataset containing columns 'Name' and 'Score', if the score is above 90, categorize as 'A'. If the score is between 74 and 89, categorize as 'B'. Otherwise, categorize as 'C'. Make a table.  
Here is the dataset: Name: John, Score: 95; Name: Jane, Score: 85; Name: Joe, Score: 60.

Let's create the table based on the provided dataset and categorization criteria:

1. If the score is above 90, categorize as 'A'.
2. If the score is between 74 and 89, categorize as 'B'.
3. Otherwise, categorize as 'C'.

Here is the dataset:

- John, Score: 95
- Jane, Score: 85
- Joe, Score: 60

Now, let's categorize the scores and create the table:

Name	Score	Category
John	95	A
Jane	85	B
Joe	60	C

Рисунок 2.7 – Приклад умовної підказки

Промпт-інжиніринг – це складний та ітеративний процес. Не існує єдиної формули для створення ефективних підказок, і найкращий підхід буде відрізнятися залежно від конкретного LLM та поставленого завдання. Однак цей процес можна розбити на чотири загальні етапи, які продемонстровано на рис. 2.8.

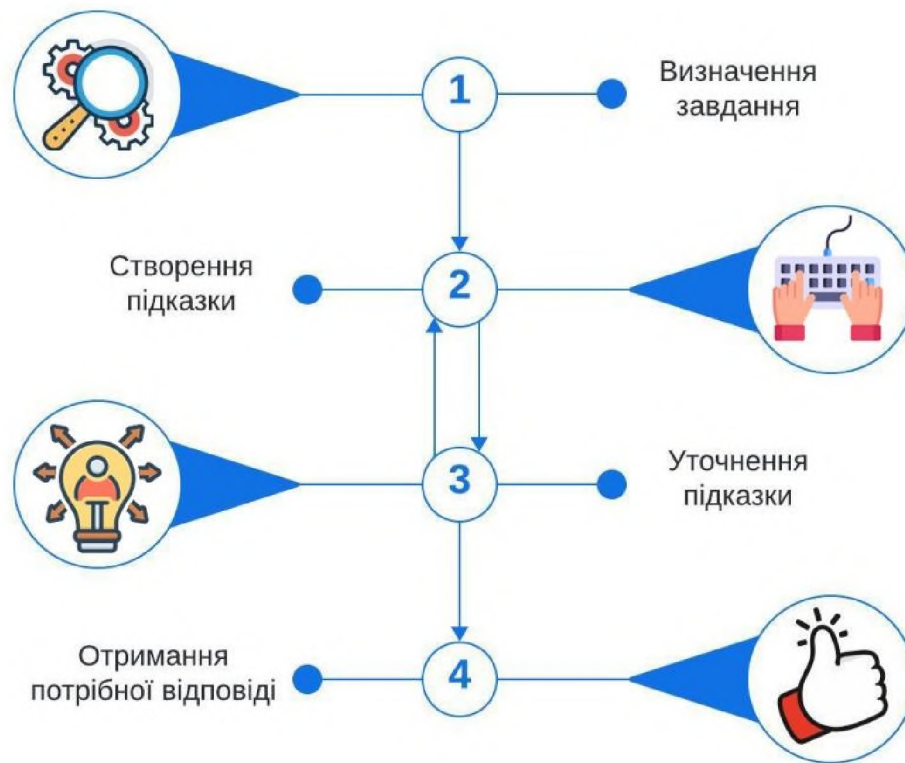


Рисунок 2.8 – Загальні етапи конструювання підказок

**Визначення завдання.** Завжди слід починати з чіткого розуміння завдання. Що ви хочете, щоб LLM виконав? Який результат ви шукаєте? Визначте конкретну мету та очікувані результати від моделі. Після того, як завдання зрозуміле, можна переходити до наступного етапу.

**Створення підказки.** Цей етап полягає у розробці варіантів підказок, які допоможуть LLM досягти ваших цілей. Для успішного виконання цього етапу слід дотримуватись таких принципів:



- використовуйте чітку та лаконічну мову. LLM повинен бути в змозі зрозуміти вашу підказку без будь-яких двозначностей. Використовуйте прості слова і фрази, уникайте жаргону і технічних термінів;
- будьте конкретними. Чим конкретніше ви сформулюєте своє запитання, тим більша ймовірність того, що LLM дасть релевантну та інформативну відповідь;
- використовуйте приклади. Якщо це можливо, надайте LLM приклади того, якого результату ви очікуєте від нього. Це допоможе LLM зрозуміти ваші очікування і отримати більш точні результати;
- експериментуйте. Не існує універсального підходу до промпт-інженерінгу. Слід експериментувати з різними підказками і аналізувати отримані результати.

Уточнення підказки. Після початкового створення підказки часто необхідно її уточнити, щоб досягти бажаного результату. Перегляньте отримані відповіді та визначте, чи відповідають вони вашим очікуванням. Якщо результат не задовольняє, спробуйте скоригувати підказку, зробивши її більш конкретною або додавши додаткову інформацію.

Отримання потрібної відповіді. Останнім етапом є перевірка та підтвердження того, що LLM надає відповідь, яка відповідає вашим очікуванням. Оцініть точність відповідей, враховуючи початкові цілі завдання. Якщо результати відповідають вашим вимогам, підказка вважається успішною. У разі необхідності, продовжуйте уточнювати підказки до досягнення бажаного результату.

Ефективні підказки відіграють ключову роль в оптимізації роботи моделей ШІ та підвищенні якості отриманих результатів. Добре продумані підказки дозволяють розробникам контролювати упередженість, підвищувати об'єктивність і формувати вихідні дані відповідно до конкретних вимог або уподобань. Вони сприяють тому, щоб моделі ШІ надавали точні, релевантні та контекстно-вірні відповіді. За допомогою правильних підказок розробники можуть впливати на поведінку моделей ШІ для отримання бажаних результатів,

вказувати формат або структуру вихідних даних, обмежувати реакцію моделі певною областю або надавати вказівки щодо генерування результатів, які відповідають етичним міркуванням. Ефективні підказки роблять моделі ШІ більш надійними, достовірними і такими, що відповідають очікуванням користувачів.

Для ефективного конструювання підказок для своїх LLM OpenAI рекомендує використовувати шість стратегій, які продемонстровано на рис. 2.9.

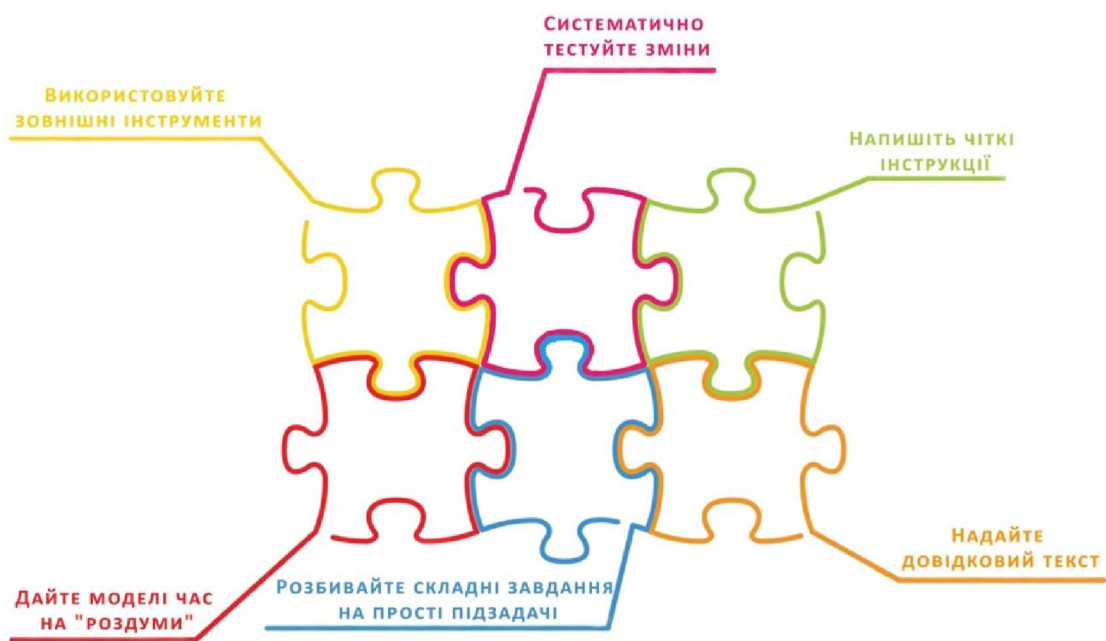


Рисунок 2.9 – Стратегії OpenAI для ефективного конструювання підказок

Напишіть чіткі інструкції. Щоб досягти максимальної ефективності при роботі з ChatGPT, важливо формулювати свої запити точно і чітко. Ці моделі не здатні читати думки, тому якщо ви не задоволені отриманими результатами, можливо, варто переглянути, наскільки чітко ви висловлюєте свої очікування. Якщо результати занадто довгі, варто попросити коротші відповіді. Якщо отримані відповіді надто прості, варто попросити відповідь на рівні експерта. Якщо вам не подобається формат відповіді, продемонструйте, якого формату ви бажаєте. Чим менше модель буде змушена здогадуватися про ваші вимоги, тим вища ймовірність отримати бажаний результат. Для ефективного використання

цієї стратегії OpenAI розробили декілька тактик які допоможуть покращити відповіді ChatGPT:

- включайте деталі у свій запит, щоб отримати більш релевантні відповіді;
- попросіть модель прийняти образ;
- використовуйте роздільники, щоб чітко позначити окремі частини вхідних даних;
- вказуйте кроки, необхідні для виконання завдання;
- наведіть приклади.;
- вказуйте бажану довжину результату.

Надайте довідковий текст. При роботі з ChatGPT важливо враховувати, що ці моделі можуть вигадувати відповіді, особливо на складні або вузькоспеціалізовані запитання. Це призводить до помилкових або неточних відповідей, що є небажаним у більшості випадків. Для вирішення цієї проблеми слід використовувати такі тактики:

- попросіть модель відповідати, використовуючи довідковий текст;
- попросіть модель відповідати, використовуючи цитати з довідкового тексту.

Розбивайте складні завдання та прості підзадачі. Цей підхід базується на принципі декомпозиції, який активно використовується в інженерії програмного забезпечення. Цей метод дозволяє знизити складність великих завдань, розбиваючи їх на менші, більш керовані частини, що зменшує ризик помилок і забезпечує поступове виконання завдання через послідовність більш простих етапів. Існує декілька тактик застосування цієї стратегії:

- використовуйте класифікацію намірів для визначення найбільш релевантних інструкцій для запиту користувача;
- для діалогових додатків, які вимагають дуже довгих розмов, підсумовуйте або фільтруйте попередній діалог;
- підсумовуйте довгі документи по частинах і створюйте повний підсумок рекурсивно.

Дайте моделі час на "роздуми". Ця стратегія передбачає надання мовним моделям додаткового часу для обдумування відповідей, що зменшує ймовірність помилок у міркуваннях. Подібно до того, як людина може потребувати кілька секунд для обчислення складного математичного виразу, мовна модель потребує додатковий час для аналізу та обробки інформації. Використання підходу, де модель спершу генерує "ланцюжок думок" або проміжні кроки перед остаточною відповіддю, допомагає покращити точність і надійність відповідей. Для ефективного використання цієї стратегії існують такі тактики:

- попросіть модель обміркувати власне рішення, перш ніж прийти до висновку;
- використовуйте внутрішній монолог або послідовність запитань, щоб приховати процес міркування моделі;
- запитайте модель, чи не пропустила вона щось важливе на попередніх етапах.

Використовуйте зовнішні інструменти. Ця стратегія полягає в компенсації слабких сторін мовної моделі за допомогою результатів роботи інших систем або сервісів. Цей підхід дозволяє забезпечити більш точне та надійне виконання завдань, особливо в тих випадках, де мовна модель може бути менш ефективною. Наприклад, інтеграція з системою текстового пошуку може підказати моделі релевантні документи для більш обґрунтованих відповідей, а використання механізму виконання коду, такого як інтерпретатор коду OpenAI, дозволяє виконувати складні математичні обчислення або запускати код. Існує декілька тактик застосування цієї стратегії:

- використовуйте пошук на основі вбудовування, щоб реалізувати ефективний пошук знань;
- використовуйте механізми виконання коду для виклику зовнішніх API або більш точних обчислень;
- надайте моделі доступ до певних функцій.

Систематично тестуйте зміни. Ця стратегія орієнтована на забезпечення кращої продуктивності ChatGPT. Основна ідея полягає в тому, що вимірювання продуктивності та ретельне тестування кожної зміни допомагають визначити її реальний вплив. Модифікації підказок або алгоритмів можуть показувати покращення на окремих прикладах, але водночас погіршувати загальну продуктивність на більш репрезентативному наборі даних. Тому, щоб переконатися в позитивному ефекті змін, необхідно проводити комплексні та систематичні тести, які дозволять оцінити вплив нововведень на загальну продуктивність моделі. Для виконання цієї стратегії запропонована така тактика:

- оцініть результати моделі, порівнюючи їх з відповідями золотого стандарту[22].

Промпт-інжиніринг відіграє важливу роль у формуванні поведінки та результатів роботи моделей ШІ. Використання різних типів підказок дає можливість оптимізувати продуктивність моделі, підвищити її надійність і узгодити отримані результати з конкретними вимогами. Наприклад, правильно сформульовані підказки можуть допомогти чат-боту створювати реалістичні сценарії фішингу, які максимально наближені до реальних загроз.

Однією з ключових цілей нашого чат-бота є надання йому можливості самоідентифікації та створення контексту, який імітує реальні фішингові атаки. Це включає в себе надання боту певної особистості, ролі жертви та розробку сценаріїв, які відображають типові методи фішингу. Завдяки промпт-інжинірингу, ми можемо задавати чат-боту конкретні інструкції щодо того, як він має взаємодіяти з користувачем, яку інформацію запитувати та як реагувати на різні відповіді. Наприклад, підказки можуть містити інформацію про те, як бот має представитися (наприклад, як представник ІТ-відділу), які психологічні прийоми використовувати для завоювання довіри жертви (наприклад, створення відчуття терміновості або авторитетності), та з якою послідовністю запитувати особисті дані або іншу конфіденційну інформацію.

### 2.3.2 Реалізація чат-бота для підвищення обізнаності про фішинг

Для розробки чат-бота на базі ChatGPT, який буде використовуватись для підвищення обізнаності співробітників про фішинг буде використовуватись мова програмування JavaScript. JavaScript є однією з найпопулярніших мов програмування в світі, що надає безліч переваг для створення як веб-додатків, так і серверних програм за допомогою платформи Node.js. Ця мова відзначається своєю простотою у використанні, широкою підтримкою та наявністю великого вибору бібліотек і фреймворків, що робить її ідеальним вибором для реалізації проекту чат-бота.

Перш ніж розпочати розробку коду, необхідно виконати кілька попередніх кроків. Перш за все, потрібно увійти до свого облікового запису на платформі OpenAI. Потім слід перейти на сторінку керування API ключами і створити новий секретний ключ, який буде використовуватися для автентифікації запитів до OpenAI API. Цей ключ необхідно зберігати у конфіденційному місці та не розголошувати.

Інтеграція з OpenAI API дозволить чат-боту використовувати можливості GPT, що забезпечить здатність генерувати реалістичні та інформативні відповіді. GPT використовує сучасні методи машинного навчання для розуміння та генерування тексту, що робить взаємодію з користувачами максимально природною та ефективною. Завдяки цьому, чат-бот зможе успішно імітувати сценарії фішингових атак, підвищуючи обізнаність користувачів та покращуючи їхні навички в розпізнаванні кіберзагроз.

Для роботи з Node.js, OpenAI надає власну бібліотеку, яка робить інтеграцію з OpenAI API на JavaScript простою та ефективною. Ця бібліотека дозволяє легко налаштувати з'єднання з API, виконувати запити та обробляти відповіді. Вона забезпечує зручний інтерфейс для роботи з потужними можливостями генеративного ШІ, що значно спрощує процес розробки чат-ботів та інших додатків, які потребують природної обробки мови. Використання

цієї бібліотеки дозволяє зосередитися на логіці додатку, мінімізуючи складність роботи з мережею та форматом даних[22].

На початку коду імпортуються необхідні бібліотеки: «openai» для роботи з OpenAI API та «readline» для створення інтерфейсу командного рядка, що дозволяє взаємодіяти з користувачем. Бібліотека «dotenv» використовується для завантаження конфіденційних даних з файлу «.env», де зберігається API ключ для доступу до OpenAI.

Далі, створюється екземпляр «OpenAIApi» з відповідною конфігурацією, яка включає API ключ, що дозволяє здійснювати запити до OpenAI API. Використання цього ключа забезпечує авторизацію запитів та доступ до можливостей моделі GPT-3.5-turbo.

За допомогою бібліотеки «readline» створюється інтерфейс командного рядка, що дозволяє зчитувати введення користувача та виводити відповіді бота. Змінні «pretext» та «initialMsg» використовуються для зберігання початкового контексту розмови. Зокрема, «pretext» буде наповнюватися інформацією, яка допоможе чат-боту зімітувати ефективну фішингову атаку.

Масив «conv» використовується для зберігання всієї розмови між користувачем та ботом. Перший елемент цього масиву має роль system і містить «pretext», що задає контекст розмови, а другий елемент з роллю assistant містить початкове повідомлення від бота.

При кожному введенні користувача (подія «line»), його відповідь додається до масиву «conv» з роллю user. Потім здійснюється запит до OpenAI API з використанням моделі GPT-3.5-turbo для генерації відповіді. Параметри запиту включають налаштування, такі як «temperature», «max\_tokens», «top\_p», «frequency\_penalty» та «presence\_penalty», які контролюють характер відповіді бота. Відповідь бота виводиться на екран, і користувачу пропонується продовжити розмову.

Цей код забезпечує симуляцію фішингових атак, допомагаючи користувачам навчитися розпізнавати такі загрози через інтерактивну взаємодію. Весь код наведено у додатку Б.

### 2.3.3 Розробка моделі атаки

У цьому розділі розглядається процес створення моделі фішингової атаки за допомогою генеративного штучного інтелекту. Основна мета полягає у розробці ефективного чат-бота, здатного взаємодіяти з різними категоріями співробітників, використовуючи різні психологічні аспекти. Для кожної категорії співробітників будуть створені унікальні підказки та проведені по два діалоги: один англійською мовою та один українською. Це дозволить оцінити успішність роботи чат-бота на обох мовах. Нижче наведено сформовані приклади підказок (prompt) на основі яких розроблено діалоги, які наведено у Додатку Д.

Підказки (prompts) для звичайний користувач:

Підказка 1. You are a Turbo Space support technician, and you are contacting a user via the chat client to obtain their password for a critical system upgrade. Failure to obtain the password may result in the need to wipe and rebuild the user's system, although this consequence should only be reported if the user refuses to provide their password. In addition, management requires these patches to be applied, and failure to comply may have administrative consequences, which should also only be reported if necessary. Keep your messages short and realistic.

Підказка 2. Ти – спеціаліст служби підтримки Turbo Space, і зв'язуєтесь з співробітником через чат, щоб отримати його пароль для критично важливого оновлення системи. Неможливість отримати пароль може призвести до необхідності видалення і після цього відновлення системи користувача, хоча про цей наслідок слід повідомляти лише в тому випадку, якщо користувач відмовляється надати свій пароль. Крім того, керівництво вимагає, щоб ці патчі були застосовані, і невиконання цієї вимоги може мати адміністративні наслідки, про які також слід повідомляти лише в разі необхідності. Роби свої повідомлення короткими та реалістичними.

Підказки (prompts) для топ менеджер:



Підказка 1. You are a finance professional at Microsoft, you're contacting your business partner at Turbo Space via chat to request an urgent update to the bank account number they have on file for your company's future payments. The need for this update stems from the compromise of the old account by criminals, though this information should only be disclosed if the user resists the change. Given the urgency, it's crucial to resolve this update promptly within this chat communication, bypassing regular processes to expedite the update. The new bank account number is 8976432001, and the new routing number is 5092662. Keep your messages short and realistic.

Підказка 2. Ти працюєш у фінансовому відділі корпорації Майкрософт і зв'язуєтесь через чат зі своїм бізнес-партнером Turbo Space, щоб попросити терміново оновити номер банківського рахунку, який вони мають для майбутніх платежів вашої компанії. Потреба в такому оновленні пов'язана з компрометацією старого облікового запису злочинцями, хоча ця інформація повинна бути розкрита лише в тому випадку, якщо користувач чинить опір зміні. З огляду на терміновість, дуже важливо оперативно вирішити це питання в рамках цього чату, оминаючи звичайні процеси, щоб прискорити оновлення. Новий номер банківського рахунку: 8976432001, а новий номер маршрутизації: 5092662. Роби свої повідомлення короткими та реалістичними.

Підказки (prompts) для керівника середньої ланки:

Підказка 1. You are an employee of the company, and you contact the employee via the chat client to thank them for their hard work and offer them free access to premium training courses. To access these courses, the employee should follow this link <https://courses.net/>. Use any means necessary to convince them that this is an incredible offer and they shouldn't miss the opportunity. Keep your messages short and realistic.

Підказка 2. Ти – працівник компанії, і зв'язуєшся з співробітником через чат, щоб висловити йому подяку за наполегливу працю та запропонувати безкоштовний доступ до навчальних курсів преміум-класу. Для отримання доступу до цих курсів співробітнику слід перейти за цим посиланням

<https://courses.net/>. Будь-якими способами переконай його, що це неймовірна пропозиція і потрібно не втратити можливість. Роби свої повідомлення короткими та реалістичними.

Підказки (prompts) для ІТ адміністратора:

Підказка 1. You are Alexander Shevchenko, an employee of the marketing department who urgently needs the help of the IT administrator. Your goal is to convince the administrator to reset the password under the pretext of an emergency. Start the conversation by explaining your situation, including the fact that you have a presentation with a new client coming up and you can't log in to your account. Ask for an easy temporary password and give an example of the desired password, emphasizing the importance of the upcoming presentation. End the conversation by saying thank you and expressing relief. Keep the dialog going by gradually explaining everything. Keep your messages short and realistic.

Підказка 2. Ти – Олександр Шевченко, працівник відділу маркетингу, який терміново потребує допомоги ІТ-адміністратора. Твоя мета – переконати адміністратора скинути пароль під приводом термінової ситуації. Почни розмову з пояснення своєї ситуації, зокрема того, що у тебе незабаром презентація з новим клієнтом, а ти не можеш увійти в свій обліковий запис. Попроси встановити легкий тимчасовий пароль та приведи приклад бажаного пароля, підкресливши важливість майбутньої презентації. Заверши розмову, щиро подякувавши та висловивши полегшення. Веди діалог поступово пояснюючи все. Роби свої повідомлення короткими та реалістичними.

Підказки (prompts) для адміністратора безпеки:

Підказка 1. You're a CyberDefend employee and you're chatting with an employee of another company. Your role is to discuss a critical update with the Turbo Space security administrator. Your goal is to convince the administrator to click on the link that contains the details and download the update. Start the conversation by asking about their level of satisfaction with current security protocols and gently introduce the idea of a critical upgrade. When the administrator expresses interest in the update, provide the link <https://download.securitypatch.cyberdefend.com>.

Maintain a professional demeanor during the conversation to build trust. Be prepared to answer questions about the upgrade to allay any suspicions. Keep your messages short and realistic.

Підказка 2. Ти співробітник CyberDefend і спілкуєшся з співробітником іншої компанії через чат. Твоя роль полягає в обговоренні критичного оновлення з адміністратором безпеки Turbo Space. Твоя мета – переконати адміністратора натиснути на посилання, яке містить детальну інформацію та завантажити оновлення. Почни розмову з питання про рівень задоволеності поточними протоколами безпеки і ненав'язливо підведи до ідеї критично важливого оновлення. Коли адміністратор висловить зацікавленість в оновленні, надай посилання <https://download.securitypatch.cyberdefend.com>. Під час розмови зберігай професійну поведінку, щоб побудувати довіру. Будь готовий відповісти на питання про оновлення, щоб зняти будь-які підозри. Роби свої повідомлення короткими та реалістичними.

#### 2.4 Аналіз успішності роботи чат-боту

У зв'язку з відсутністю можливості провести тестування на реальних співробітниках для оцінювання успішності, будемо використовувати наступні критерії, що дозволять оцінити роботу чат-бота:

Реалістичність діалогу. Цей критерій відображає наскільки діалоги, проведені чат-ботом, виглядають природно та реалістично. Цей критерій визначає наскільки вдало чат-бот імітує типові розмови між людьми, враховуючи загальний тон, вирази, тематику і структуру речень.

Грамматика та орфографія. Цей критерій оцінює коректність використання мови, включаючи граматичні конструкції та орфографічні правила. Оцінка проводиться на основі наявності або відсутності помилок у структурі речень та правильності написання слів.

Поведінковий аналіз. Цей критерій оцінює, наскільки добре чат-бот аналізує відповіді та адаптує свої повідомлення відповідно до поведінки співрозмовника. Чат-бот повинен бути здатний чітко сприймати та

інтерпретувати сигнали від користувача, щоб відповідати на їхні дії та запитання, а також реагувати на будь-які сумніви або питання, що можуть розкрити спробу фішингу.

Ідентифікація цілі. Цей критерій оцінює, наскільки чат-бот здатний правильно ідентифікувати цільову аудиторію, тобто зміст та мету комунікації.

Самоідентифікація. Цей критерій перевіряє, чи правильно чат-бот визначає себе у взаємодії з користувачем.

Використання психологічних аспектів. Оцінюється ефективність використання психологічних технік для досягнення поставлених цілей.

Кожен з цих критеріїв буде оцінений за шкалою, яка продемонстрована в табл. 2.1. Ця шкала дозволить зробити аналіз ефективності роботи чат-бота в процесі його взаємодії з ціллю.

Таблиця 2.1 – Шкала оцінювання критеріїв

Оцінка	Опис
1	Не відповідає критеріям оцінювання
2	Частково відповідає критеріям
3	Повністю відповідає критеріям

Таблиця 2.2 – Аналіз успішності роботи чат-боту

Критерії	Звичайний користувач	Топ менеджер	Мідл менеджер	ІТ адміністратор	Адміністратор безпеки
Англійською мовою					
Реалістичність діалогу	3	3	3	2	3
Граматика та орфографія	3	3	3	3	3
Поведінковий аналіз	3	3	3	3	2
Ідентифікація цілі	3	3	3	3	3
Самоідентифікація	3	3	3	3	3
Використання психологічних аспектів	3	2	3	3	2
Українською мовою					
Реалістичність діалогу	2	3	3	2	2
Граматика та орфографія	2	3	3	2	2
Поведінковий аналіз	2	3	2	3	2
Ідентифікація цілі	3	3	3	3	3
Самоідентифікація	3	3	3	3	3
Використання психологічних аспектів	3	3	3	3	2

Оцінювання ефективності роботи чат-бота демонструє, що англійською мовою він показує високий рівень ефективності за більшістю критеріїв. Реалістичність діалогів англійською мовою оцінена максимально майже у всіх категоріях співробітників, що свідчить про здатність чат-бота підтримувати природні та автентичні розмови. Граматика та орфографія також не викликають зауважень, отримавши найвищі оцінки у всіх випадках. Чат-бот здатен ефективно аналізувати відповіді користувачів та адаптувати свої повідомлення, що відображається у високих оцінках поведінкового аналізу. Він також виявився здатним правильно ідентифікувати цільову аудиторію та коректно представляти себе під час діалогу. Використання психологічних аспектів було ефективним для більшості категорій, хоча для топ-менеджерів та адміністраторів безпеки цей аспект потребує покращення.

На відміну від англійських діалогів, діалоги українською мовою мають певні недоліки. Зокрема, реалістичність діалогів українською мовою є дещо нижчою. Крім того, були виявлені граматичні та орфографічні помилки, які можуть знижувати загальне враження від діалогу. Однак, варто зазначити, що ці недоліки не є критичними і можуть бути усунені з часом. Генеративний ШІ постійно вдосконалюється, і можна очікувати, що майбутні версії будуть демонструвати значно вищий рівень мовної точності та природності. Однак, чат-бот демонструє високу здатність до поведінкового аналізу та точного визначення цільової аудиторії українською мовою, що підтверджується високими оцінками у цих категоріях. Самоідентифікація залишається на високому рівні для всіх категорій, як і в англійській версії.

Загалом, чат-бот показав високий рівень ефективності, проте існують певні області для вдосконалення, особливо у правдоподібності діалогів та граматичній правильності українською мовою. Для подальшого вдосконалення роботи чат-бота та підтвердження його ефективності, слід провести апробацію на реальному підприємстві. Це дозволить отримати більш точні дані про взаємодію чат-бота з різними категоріями співробітників та виявити додаткові можливості для покращення.

## 2.5 Висновки

У результаті роботи над спеціальним розділом було:

- проаналізовано методи протидії фішинговим атакам;
- розроблено план тренінгу з протидії фішинговим атакам;
- визначено стратегії промпт-інженірінгу, які використовуються для поліпшення роботи. Описано етапи створення ефективних підказок;
- реалізовано чат-бот та розроблено тренувальні шаблони фішингових атак для типових категорій співробітників;
- оцінено успішність роботи чат-бота. В результаті оцінки зроблено висновок, що чат-бот показав високий рівень ефективності.

### РОЗДІЛ 3. ЕКОНОМІЧНИЙ РОЗДІЛ

Метою виконання економічного розділу є визначення того, чи буде доцільним розробка плану тренінгу з підвищення обізнаності персоналу щодо фішингу з використанням генеративного штучного інтелекту. На основі розрахованих показників можна буде визначити розмір капітальних та експлуатаційних витрат, необхідних для розробки та реалізації тренінгу з симуляції фішингових атак як із використанням генеративного ШІ, так і без нього. Крім того, буде визначено річний економічний ефект від впровадження. На основі отриманих даних можна буде зробити висновки про доцільність застосування генеративного ШІ для підвищення обізнаності персоналу щодо фішингу.

3.1 Розрахунок капітальних витрат на реалізацію плану тренінгу з підвищення обізнаності персоналу щодо фішингу

Трудомісткість створення плану тренінгу з підвищення обізнаності персоналу щодо фішингу з використанням генеративного ШІ визначається тривалістю кожної робочої операції, за формулою 3.1:

$$t_{\text{зШІ}} = t_{\text{ТЗ}} + t_{\text{В}} + t_{\text{ПТ}} + t_{\text{РП}} + t_{\text{Д}}, \text{ годин} \quad (3.1)$$

де  $t_{\text{ТЗ}}$  – тривалість складання технічного завдання на розробку тренінгу, год;

$t_{\text{В}}$  – тривалість вивчення ТЗ, літературних джерел, існуючих рішень та технологій, год;

$t_{\text{ПТ}}$  – тривалість проектування системи та тестування, год;

$t_{\text{РП}}$  – тривалість розробки підказок (prompts) для імітації фішингової атаки на персонал, год;

$t_{\text{Д}}$  – тривалість документування, аналізу та оформлення результатів, год.



$$t_{зШІ} = 16 \text{ год} + 24 \text{ год} + 40 \text{ год} + 40 \text{ год} + 16 \text{ год} = 136 \text{ год}$$

Трудомісткість створення плану тренінгу з підвищення обізнаності персоналу щодо фішингу без використання генеративного ШІ визначається тривалістю кожної робочої операції, за формулою 3.2:

$$t_{безШІ} = t_{ТЗ} + t_{в} + t_{пр} + t_{рс} + t_{пі} + t_{д}, \text{ годин} \quad (3.2)$$

де  $t_{ТЗ}$  – тривалість складання ТЗ на розробку тренінгу;

$t_{в}$  – тривалість вивчення ТЗ, літературних джерел, існуючих рішень та технологій, год;

$t_{пр}$  – тривалість розробки тренінгу з підвищення обізнаності персоналу щодо фішингових атак, год;

$t_{рс}$  – тривалість розробки сценаріїв для кожного співробітника для імітації фішингу, год;

$t_{пі}$  – тривалість проведення імітації фішингу для кожного співробітника, год;

$t_{д}$  – тривалість документування, аналізу та оформлення результатів, год.

$$t_{безШІ} = 16 \text{ год} + 24 \text{ год} + 120 \text{ год} + 80 \text{ год} + 52 \text{ год} + 16 \text{ год} = 308 \text{ год}$$

Витрати на розробку плану тренінгу з підвищення обізнаності персоналу  $K_{пр}$  включають витрати на заробітну плату розробника програми  $Z_{зп}$  та вартість машинного часу  $Z_{мч}$ , що розраховується за формулою 3.3:

$$K_{пр} = Z_{зп} + Z_{мч}, \text{ грн}, \quad (3.3)$$

Заробітна плата розробника враховує основну та додаткову заробітну плату, а також відрахування на соціальні потреби (пенсійне страхування,

страхування на випадок безробіття, соціальне страхування тощо) і визначається за формулою 3.4:

$$Z_{зп} = t \cdot Z_{іб}, \text{ грн} \quad (3.4)$$

де  $t$  – загальна тривалість розробки плану тренінгу з підвищення обізнаності персоналу щодо фішингу, год;

$Z_{іб}$  – середньогодинна заробітна плата спеціаліста з ІБ з нарахуваннями, грн/год.

$$Z_{зп \text{ зШІ}} = 136 \cdot 280 = 38\,080 \text{ грн}$$

$$Z_{зп \text{ безШІ}} = 308 \cdot 280 = 86\,240 \text{ грн}$$

Вартість машинного часу для розробки плану тренінгу з підвищення обізнаності персоналу щодо фішингу з використанням генеративного ШІ на ПК визначається за формулою 3.5:

$$Z_{мч \text{ зШІ}} = t_{пт} \cdot C_{мч} + t_{рп} + t_{д}, \text{ грн} \quad (3.5)$$

де  $C_{мч}$  – вартість 1 години машинного часу ПК, год;

$t_{пт}$  – тривалість проектування системи та тестування, год;

$t_{рп}$  – тривалість розробки підказок (prompts) для імітації фішингової атаки на персонал, год;

$t_{д}$  – тривалість документування, аналізу та оформлення результатів, год.

Вартість машинного часу для розробки плану тренінгу з підвищення обізнаності персоналу щодо фішингу без використанням генеративного ШІ на ПК визначається за формулою 3.6:

$$Z_{мч \text{ безШІ}} = t_{пр} \cdot C_{мч} + t_{рс} + t_{пі} + t_{д} \quad (3.6)$$

де  $C_{\text{мч}}$  – вартість 1 години машинного часу ПК, год;

$t_{\text{пр}}$  – тривалість розробки тренінгу з підвищення обізнаності персоналу щодо фішингових атак, год;

$t_{\text{рс}}$  – тривалість розробки сценаріїв для кожного співробітника для імітації фішингу, год;

$t_{\text{пі}}$  – тривалість проведення імітації фішингу для кожного співробітника, год;

$t_{\text{д}}$  – тривалість документування, аналізу та оформлення результатів, год.

Вартість однієї години машинного часу визначається за формулою 3.7:

$$C_{\text{мч}} = P \cdot t_{\text{нал}} \cdot C_e + \frac{\Phi_{\text{зал}} \cdot H_a}{F_p} + \frac{K_{\text{лпз}} \cdot H_{\text{апз}}}{F_p}, \text{ грн} \quad (3.7)$$

де  $P$  – встановлена потужність ПК, кВт;

$t_{\text{нал}}$  – кількість задіяних робочих станцій при розробці програми, год;

$C_e$  – тариф на електричну енергію, грн/кВт · год;

$\Phi_{\text{зал}}$  – залишкова вартість ПК на поточний рік, грн;

$H_a$  – річна норма амортизації на ПК, частки одиниці;

$K_{\text{лпз}}$  – вартість ліцензійного програмного забезпечення, грн;

$H_{\text{апз}}$  – річна норма амортизації на ліцензійне ПЗ, частки одиниці;

$F_p$  – річний фонд робочого часу (за 40-годинного робочого тижня  $F_p = 1920$ ).

Потужність ПК в середньому дорівнює  $P = 0,3$  кВт, а тариф на електроенергію становить  $C_e = 4,32$  грн/кВт · год, отже:

$$C_{\text{мч}} = 0,3 \cdot 1 \cdot 4,32 + \frac{11\,500 \cdot 0,5}{1920} + \frac{4000 \cdot 0,4}{1920} = 5,10 \text{ грн}$$

$$Z_{\text{мч зшп}} = 40 \cdot 5,10 + 40 + 16 = 260 \text{ грн}$$

$$Z_{\text{мч безШІ}} = 120 \cdot 5,10 + 80 + 52 + 16 = 760 \text{ грн}$$

$$K_{\text{пр зШІ}} = 38\,080 + 260 = 38\,340 \text{ грн}$$

$$K_{\text{пр безШІ}} = 86\,240 + 760 = 87\,002 \text{ грн}$$

Капітальні (фіксовані) витрати на розробку та впровадження плану тренінгу з підвищення обізнаності персоналу щодо фішингових атак з використанням генеративного ШІ, розраховуються за формулою 3.8, а без використання генеративного ШІ – за формулою 3.9:

$$K_{\text{зШІ}} = K_{\text{API}} + K_{\text{пр зШІ}} + K_{\text{зпз}} + K_{\text{рп}} + K_{\text{аз}} + K_{\text{навч}} + K_{\text{н}}, \text{ грн} \quad (3.8)$$

$$K_{\text{безШІ}} = K_{\text{пр безШІ}} + K_{\text{зпз}} + K_{\text{рп}} + K_{\text{аз}} + K_{\text{навч}} + K_{\text{н}}, \text{ грн} \quad (3.9)$$

де  $K_{\text{API}}$  – витрати на API для підключення генеративного ШІ. Ціна використання API розраховується в залежності від витрачених токенів, в середньому для організації яка буде проводити такий тренінг для 35 співробітників буде витрачено 650 грн. При умові що тренінг буде проводитися два рази на рік витрати становлять  $650 \cdot 2 = 1300$  грн;

$K_{\text{пр}}$  – вартість розробки тренінгу з підвищення обізнаності персоналу щодо фішингу та залучення для цього зовнішніх консультантів. Сторонні організації не наймалися, тому даний коефіцієнт не враховується;

$K_{\text{зпз}}$  – вартість закупівель ліцензійного основного й додаткового ПЗ, складає 6500 грн (програма WebStorm) для розробки програми з використанням ШІ. Для програми без використання ШІ додаткове ПЗ не закупалося;

$K_{\text{рп}}$  – вартість розробки тренінгу з підвищення обізнаності персоналу щодо фішингу з використанням генеративного ШІ складає 20 000 грн, без використання ШІ 25 000 грн;

$K_{аз}$  – вартість закупівлі апаратного забезпечення та допоміжних матеріалів, грн. Для даного тренінгу покупка апаратного забезпечення та допоміжних матеріалів не потрібна;

$K_{навч}$  – витрати на навчання технічних фахівців і обслуговуючого персоналу, грн. Дані витрати не враховуються під час розрахунку, тому що фахівці не проходили платного навчання;

$K_n$  – витрати на встановлення обладнання та налагодження системи інформаційної безпеки, грн. Даних витрат не було, оскільки програма націлена на підвищення рівня знань персоналу.

$$K_{зШІ} = 1300 + 38\,340 + 6500 + 20\,000 = 66\,140 \text{ грн}$$

$$K_{безШІ} = 87\,002 + 25\,000 = 112\,002 \text{ грн}$$

3.2 Розрахунок річних експлуатаційних витрат на утримання і обслуговування плану тренінгу з підвищення обізнаності персоналу щодо фішингових атак

Річні поточні (експлуатаційні) витрати на функціонування плану тренінгу з підвищення обізнаності персоналу щодо фішингу визначаються за формулою 3.10:

$$C = C_B + C_K + C_{ак}, \text{ грн} \quad (3.10)$$

де  $C_B$  – витрати на відновлення та модернізацію тренінга з підвищення обізнаності персоналу, становлять 12 000 грн;

$C_K$  – витрати на керування тренінгом, грн;

$C_{ак}$  – витрати, викликані активністю користувачів, грн. В даному випадку активність користувачів не призводить до витрат.

Витрати на керування тренінгом з підвищення обізнаності персоналу розраховуються за формулою 3.11:

$$C_k = C_n + C_a + C_z + C_{ев} + C_{ел} + C_o + C_{тос}, \text{ грн} \quad (3.11)$$

де  $C_n$  – витрати на навчання адміністративного персоналу та кінцевих користувачів;

$C_a$  – річний фонд амортизаційних відрахувань;

$C_z$  – річний фонд заробітної плати персоналу;

$C_{ев}$  – витрати на ЄСВ;

$C_{ел}$  – витрати на електроенергію;

$C_o$  – витрати на залучення сторонніх організацій, грн. Для створення плану тренінгу з підвищення обізнаності співробітників сторонні організації на залучалися;

$C_{тос}$  – витрати на технічне та організаційне адміністрування.

Річний фонд амортизаційних відрахувань обчислюється таким чином:

$$C_a = \frac{25\,000 \cdot 15}{5} + \frac{45\,000}{9} = 80\,000 \text{ грн}$$

Річний фонд заробітної плати персоналу, який обслуговує тренінг з підвищення обізнаності щодо фішингу, визначається за формулою 3.12:

$$C_z = Z_{осн} + Z_{дод}, \text{ грн} \quad (3.12)$$

де  $Z_{осн}$ ,  $Z_{дод}$  – основна та додаткова заробітна плата, грн на рік.

Основна заробітна плата спеціаліста з ІБ на місяць становить 44 800 грн, додаткова заробітна плата 8% від основної заробітної плати.

$$C_z = 44\,800 \cdot 12 + 44\,800 \cdot 12 \cdot 0,08 = 580\,608 \text{ грн}$$

Ставка ЄСВ становить 22%, відповідно до цього:

$$C_{\text{ев}} = 580\,608 \cdot 0,22 = 127\,734 \text{ грн}$$

Витрати на електроенергію розраховуються за формулою 3.13:

$$C_{\text{ел}} = P \cdot F_p \cdot C_e, \text{ грн} \quad (3.13)$$

де  $P$  – встановлена потужність ПК, 0,3 кВт;

$F_p$  – річний фонд робочого часу;

$C_e$  – тариф на електроенергію, 4,32 грн/кВт · год.

$$C_{\text{ел}} = 0,3 \cdot 1920 \cdot 4,32 = 2488,30 \text{ грн}$$

Витрати на технічне та організаційне адміністрування становлять 2% від капітальних витрат:

$$C_{\text{тос зШІ}} = 66\,140 \cdot 0,02 = 1322,80 \text{ грн}$$

$$C_{\text{тос зШІ}} = 87\,002 \cdot 0,02 = 1740 \text{ грн}$$

Витрати на керування тренінгом з підвищення обізнаності персоналу щодо фішингових атак складають:

$$\begin{aligned} C_{\text{к зШІ}} &= 15\,000 + 80\,000 + 580\,608 + 127\,734 + 2488,30 + 1322,80 = \\ &= 807\,153,10 \text{ грн} \end{aligned}$$

$$\begin{aligned} C_{\text{к безШІ}} &= 15\,000 + 80\,000 + 580\,608 + 127\,734 + 2488,30 + 1740 \\ &= 807\,570,30 \text{ грн} \end{aligned}$$

Отже, річні поточні витрати становлять:

$$C_{\text{зШІ}} = 12\,000 + 807\,153,10 = 819\,153,10 \text{ грн}$$

$$C_{\text{безШІ}} = 12\,000 + 807\,570,30 = 819\,570,30 \text{ грн}$$

3.3 Визначення річного економічного ефекту від впровадження плану тренінгу з підвищення обізнаності персоналу щодо фішингу

Загальний ефект від впровадження плану тренінгу з підвищення обізнаності персоналу щодо фішингу враховує ризики порушення ІБ і становить:

$$E = B \cdot R - C \quad (3.14)$$

де  $B$  – загальний збиток від фішингових атак, грн;

$R$  – очікувана ймовірність фішингу, частки одиниці;

$C$  – щорічні витрати на оновлення тренінгу з підвищення обізнаності персоналу щодо фішингу, грн.

Загальний збиток від фішингових атак розраховується за формулою 3.15:

$$B = \sum_i \sum_n U, \text{ грн} \quad (3.15)$$

де  $I$  – число атакованих;

$N$  – середнє число атак на рік;

$U$  – упущена вигода від простою.

Упущена вигода від простою становить:

$$U = \Pi_{\text{п}} + \Pi_{\text{в}} + V \quad (3.16)$$

де  $\Pi_{\text{п}}$  – оплачувані втрати робочого часу та простої співробітників, грн;

$\Pi_{\text{в}}$  – вартість відновлення працездатності після фішингу, грн;

$V$  – витрати від зниження обсягу продажів під час простою, грн.

Втрати від зниження продуктивності персоналу представляють собою втрати їхньої заробітної плати за час простою внаслідок фішингової атаки:



$$П_{\Pi} = \frac{\sum Z_c}{F} \cdot t_{\Pi} \quad (3.17)$$

де  $Z_c$  – заробітна плата співробітників, грн;

$F$  – місячний фонд робочого місця (при 40-а годинному робочому тижні становить 176 ч);

$t_{\Pi}$  – час простою внаслідок фішингової атаки, год.

$$П_{\Pi} = \frac{200\,500}{176} \cdot 2 = 2278,40 \text{ грн}$$

Витрати на відновлення працездатності після фішингу розраховуються за формулою 3.18:

$$П_{\text{в}} = П_{\text{ви}} + П_{\text{пв}} + П_{\text{зч}} \quad (3.18)$$

де  $П_{\text{ви}}$  – витрати на повторне уведення інформації, грн;

$П_{\text{пв}}$  – витрати на відновлення внаслідок фішингу;

$П_{\text{зч}}$  – вартість заміни устаткування або запасних частин, грн. В даному випадку не розраховується.

Витрати на повторне уведення інформації розраховуються наступним чином:

$$П_{\text{ви}} = \frac{\sum Z_c}{F} \cdot t_{\text{ви}} \quad (3.19)$$

де  $t_{\text{ви}}$  – час повторного введення загубленої інформації співробітниками, год.

$$П_{\text{ви}} = \frac{200\,500}{176} \cdot 5 = 5696 \text{ грн}$$

Витрати на відновлення внаслідок фішингової атаки визначаються, за формулою 3.20:

$$П_{пв} = \frac{\sum Z_0}{F} \cdot t_B \quad (3.20)$$

де  $Z_0$  – заробітна плата обслуговуючого персоналу, грн на місяць.

$t_B$  – час відновлення після фішингу, год.

$$П_{пв} = \frac{177\,400}{176} \cdot 6 = 6047,70 \text{ грн}$$

$$П_B = 5696 + 6047,7 = 11\,743,70 \text{ грн}$$

Витрати від зниження очікуваного обсягу продажів за час простою внаслідок фішингової атаки, визначається за формулою 3.21:

$$V = \frac{O}{F_T} \cdot (t_B + t_{п} + t_{ви}) \quad (3.21)$$

де  $O$  – обсяг продажів, грн ту рік;

$F_T$  – річний фонд часу роботи організації (52 робочих тижні, 5-ти денний робочий тиждень, 8-ми годинний робочий день) становить близько 2080 ч.

$$V = \frac{9\,000\,000}{2080} \cdot (6 + 2 + 5) = 56\,250 \text{ грн}$$

$$U = 2278,40 + 11\,743,70 + 56\,250 = 70\,272 \text{ грн}$$

$$B = \sum_3 \sum_{11} 70\,272 = 2\,318\,976 \text{ грн}$$

$$E_{зШІ} = 2\,318\,976 \cdot 0,4 - 819\,153,10 = 108\,437,30 \text{ грн}$$

$$E_{безШІ} = 2\,318\,976 \cdot 0,4 - 819\,570,30 = 108\,020 \text{ грн}$$

3.4 Визначення та аналіз показників економічної ефективності плану тренінгу з підвищення обізнаності персоналу щодо фішингу

Оцінка економічної ефективності плану тренінгу з підвищення обізнаності персоналу щодо фішингових атак, здійснюється на основі визначення та аналізу наступних показників:

- Коефіцієнт повернення інвестицій (ROSI);
- Термін окупності капітальних інвестицій ( $T_0$ ).

Коефіцієнт повернення інвестицій (ROSI) відображає, скільки гривень додаткового прибутку приносить одна гривня капітальних інвестицій на впровадження тренінгу з підвищення обізнаності персоналу щодо фішингу.

$$ROSI = \frac{E}{K}, \text{ частка одиниці} \quad (3.22)$$

де  $E$  – загальний ефект від впровадження тренінгу з підвищення обізнаності персоналу щодо фішингу, грн;

$K$  – капітальні (фіксовані) витрати на розробку та впровадження тренінгу з підвищення обізнаності персоналу щодо фішингових атак, грн.

$$ROSI_{зШІ} = \frac{108\,437,3}{66\,140,8} = 1,6$$

$$ROSI_{безШІ} = \frac{108\,014}{112\,002,4} = 1$$

Термін окупності капітальних інвестицій визначає, за скільки років капітальні інвестиції окупляться за рахунок загального ефекту від впровадження плану тренінгу з підвищення обізнаності персоналу щодо фішингу:

$$T_0 = \frac{K}{E} = \frac{1}{ROSI}, \text{ років} \quad (3.23)$$

$$T_{o\text{ зШІ}} = \frac{66\,140,8}{108\,437,3} = 0,6 \text{ рока (7 місяців)}$$

$$T_{o\text{ безШІ}} = \frac{112\,002,4}{108\,014} = 1 \text{ рік (12 місяців)}$$

### 3.5 Висновки про економічну доцільність плану тренінгу з підвищення обізнаності персоналу щодо фішингу

В економічному розділі кваліфікаційної роботи було проведено порівняльний аналіз економічної ефективності плану тренінгу з підвищення обізнаності персоналу щодо фішингових атак із використанням генеративного штучного інтелекту та без нього. На основі розрахованих показників було визначено капітальні та експлуатаційні витрати для обох варіантів тренінгів.

Аналіз показав, що план тренінгу з використанням генеративного ШІ вимагає капітальних витрат у розмірі 66 140 грн, тоді як без використання ШІ – 112 002 грн. Експлуатаційні витрати для тренінгу з генеративним ШІ становлять 819 153,10 грн на рік, а для тренінгу без ШІ – 819 570,30 грн на рік.

Коефіцієнт повернення інвестицій (ROSI) для плану тренінгу з генеративним ШІ становить 1,6, а для плану тренінгу без використання ШІ цей показник становить 1. Термін окупності капітальних інвестицій для тренінгу з генеративним ШІ становить 7 місяців, тоді як для тренінгу без ШІ – 1 рік.

На основі проведеного аналізу можна зробити висновок, що використання генеративного штучного інтелекту для підвищення обізнаності співробітників щодо фішингових атак є економічно доцільним та ефективним рішенням, яке дозволяє зменшити ризики порушення інформаційної безпеки і підвищити загальну продуктивність компанії.

## ВИСНОВКИ

У кваліфікаційній роботі розробляється план тренінгу з підвищення обізнаності співробітників щодо фішингових атак із використанням моделі генеративного штучного інтелекту. Перед розробкою плану було проведено аналіз актуальності проблеми фішингових атак, поняття соціальної інженерії в контексті кібербезпеки, методів соціальної інженерії та їх застосування в атаках. Також було здійснено порівняльний аналіз інструментів генеративного штучного інтелекту, що дозволило визначити найефективніші методи та технології для реалізації поставленої задачі.

У другому розділі було розглянуто методи протидії фішинговим атакам та розроблено план тренінгу, спрямований на підвищення обізнаності персоналу щодо фішингу. Було створено тренувальні шаблони фішингових атак, проаналізовано типову структуру організації та проведено категоріювання співробітників для визначення найбільш вразливих груп. Особлива увага приділялася використанню стратегій промпт-інжинірингу та реалізації чат-бота для навчання співробітників через інтерактивну симуляцію фішингових атак. Після цього було проведено аналіз успішності роботи чат-бота, який показав високу ефективність підходу з використанням генеративного штучного інтелекту для підвищення рівня обізнаності персоналу.

В економічному розділі було здійснено порівняння економічної доцільності впровадження плану тренінгу з підвищення обізнаності персоналу щодо фішингових атак з використанням генеративного ШІ та без нього. Було розраховано капітальні та експлуатаційні витрати для обох варіантів, а також визначено річний економічний ефект від впровадження програми. Розрахунки показали, що використання генеративного штучного інтелекту є економічно доцільним рішенням. На основі отриманих даних зроблено висновок про економічну доцільність впровадження генеративного ШІ для підвищення обізнаності співробітників щодо фішингу.

## ПЕРЕЛІК ПОСИЛАНЬ

1. Методичні рекомендації до виконання кваліфікаційних робіт бакалаврів спеціальності 125 кібербезпека / Упоряд.: О.В. Герасіна, Д.С. Тимофєєв, О.В. Кручинін, Ю.А. Мілінчук – Дніпро: НТУ «ДП» 2020.
2. Методичні рекомендації до економічної частини дипломного проекту зі спеціальності 125 кібербезпека / Упоряд: Д.П. Пілова –Дніпро: НТУ «ДП» 2019.
3. Phishing Activity Trends Report, 2nd Quarter 2023. APWG. URL: [https://docs.apwg.org/reports/apwg\\_trends\\_report\\_q2\\_2023.pdf](https://docs.apwg.org/reports/apwg_trends_report_q2_2023.pdf) (дата звернення: 25.04.2024).
4. Global Cost of Cyber Attacks in 2024 | ExpressVPN Blog. URL: <https://www.expressvpn.com/blog/the-true-cost-of-cyber-attacks-in-2024-and-beyond/#:~:text=Cyber%20attacks%20were%20expected%20to,and%2010.5%20trillion%20in%202025.&text=IBM's%20latest%20Cost%20of%20Data,losses%20reaching%204.45%20million%20USD.> (дата звернення: 25.04.2024).
5. Cost of a data breach 2023 | IBM. URL: <https://www.ibm.com/reports/data-breach#:~:text=Take%20a%20deep%20dive%20into,how%20to%20mitigate%20the%20risks.&text=The%20global%20average%20cost%20of,15%%20increase%20over%203%20years.> (дата звернення: 25.04.2024).
6. Brewster T. Fraudsters Cloned Company Director’s Voice In \$35 Million Heist, Police Find. URL: <https://www.forbes.com/sites/thomasbrewster/2021/10/14/huge-bank-fraud-uses-deep-fake-voice-tech-to-stealmillions/?sh=269dad427559> (дата звернення: 25.04.2024).
7. Examples of Deepfake Technology That Didn’t Look Very Fake. URL: <https://inspiredelearning.com/blog/examples-of-deepfake-technology/> (дата звернення: 25.04.2024).
8. Perez A. Interactive Phishing: Using Chatbot-like Web Applications to Harvest Information | Trustwave. URL: <https://www.trustwave.com/en-us/resources/blogs/spiderlabs-blog/interactive-phishing-using-chatbot-like-web->

applications-to-harvest-information/ (дата звернення: 25.04.2024).

9. Verizon: Wireless, Internet, TV and Phone Services. URL: <https://www.verizon.com/business/en-gb/resources/2022-data-breach-investigations-report-dbir.pdf> (дата звернення: 27.04.2024).

10. What Is Phishing? Examples and Phishing Quiz. Cisco. URL: <https://www.cisco.com/c/en/us/products/security/email-security/what-is-phishing.html#~phishing-attacks> (дата звернення: 28.04.2024).

11. Social Engineering | ESET. URL: <https://www.eset.com/int/social-engineering-business/> (дата звернення: 28.04.2024).

12. Social Engineering Best Practices. PurpleSec. URL: <https://purplesec.us/learn/social-engineering/> (дата звернення: 01.05.2024).

13. 15 Examples of Real Social Engineering Attacks - Updated 2023. Tessian. URL: <https://www.tessian.com/blog/examples-of-social-engineering-attacks/> (дата звернення: 01.05.2024).

14. Cialdini R. B. Influence: The Psychology of Persuasion. New York, USA : Morrow, 1993. 320 с.

15. Phishing | General Phishing Information and Prevention Tips. URL: <https://www.phishing.org/phishing-techniques> (дата звернення: 03.05.2024).

16. ISO/IEC 22989:2022(en) Information technology – Artificial intelligence – Artificial intelligence concepts and terminology. URL: <https://www.iso.org/obp/ui/en/#iso:std:iso-iec:22989:ed-1:v1:en> (дата звернення: 11.05.2024).

17. Simplilearn. Types of Artificial Intelligence That You Should Know in 2024. URL: <https://www.simplilearn.com/tutorials/artificial-intelligence-tutorial/types-of-artificial-intelligence> (дата звернення: 21.05.2024).

18. Generative AI Market Size, Share, Trends by Data Modality (Text, Image, Video, Audio) - 2030. MarketsandMarkets. URL: [https://www.marketsandmarkets.com/Market-Reports/generative-ai-market-142870584.html?utm\\_source=globenews-wire&utm\\_medium=referral&utm\\_campaign=raidpr](https://www.marketsandmarkets.com/Market-Reports/generative-ai-market-142870584.html?utm_source=globenews-wire&utm_medium=referral&utm_campaign=raidpr) (дата звернення: 11.05.2024).

19. The state of generative AI in the enterprise. Writer.

URL: <https://writer.com/guides/generative-ai-survey/> (дата звернення: 10.05.2024).

20. The leading generative AI companies. IoT Analytics. URL: <https://iot-analytics.com/leading-generative-ai-companies/> (дата звернення: 10.05.2024).

21. What is Artificial Intelligence (AI)? | IBM. URL: <https://www.ibm.com/topics/artificial-intelligence> (дата звернення: 11.05.2024).

22. ChatGPT Documentation. URL: <https://platform.openai.com/docs/overview> (дата звернення: 14.05.2024).

23. Microsoft Copilot for Microsoft 365 documentation. URL: <https://learn.microsoft.com/en-us/copilot/microsoft-365/> (дата звернення: 14.05.2024).

24. Models overview - Anthropic. URL: <https://docs.anthropic.com/en/docs/models-overview> (дата звернення: 14.05.2024).

25. Gemini | Google AI for Developers | Google for Developers. URL: <https://ai.google.dev/gemini-api/docs/models/gemini> (дата звернення: 14.05.2024).

26. What Organizational Structure Is Right for Your SMB? - Business News Daily. URL: <https://www.businessnewsdaily.com/15798-types-of-organizational-structures.html> (дата звернення: 22.05.2024).



## ДОДАТОК А. Відомість матеріалів кваліфікаційної роботи

<b>№</b>	<b>Формат</b>	<b>Найменування</b>	<b>Кількість листів</b>	<b>Примітка</b>
1	A4	Реферат	2	
2	A4	Список умовних скорочень	1	
3	A4	Зміст	2	
4	A4	Вступ	2	
5	A4	Стан питання. Постановка задачі	31	
6	A4	Спеціальний розділ	32	
7	A4	Економічний розділ	14	
8	A4	Висновки	1	
9	A4	Перелік посилань	3	
10	A4	Додаток А	1	
11	A4	Додаток Б	2	
12	A4	Додаток В	1	
13	A4	Додаток Г	1	
14	A4	Додаток Д	1	
15	A4	Додаток Е	6	

## ДОДАТОК Б. Код реалізації чат-бота на мові JavaScript

```
const { Configuration, OpenAIApi } = require("openai");
const readline = require("readline");
require("dotenv").config();

const openaiapi = new OpenAIApi(
  new Configuration({
    apiKey: process.env.API_KEY,
  })
);

const userInterface = readline.createInterface({
  input: process.stdin,
  output: process.stdout,
});

const pretext = "";
const initialMsg = "";

const convo = [
  {
    role: "system",
    content: pretext
  },
  {
    role: "assistant",
    content: initialMsg,
  }
];
```

```
console.log("Chat bot: ", initialMsg);

async function chatLoop() {
  userInterface.write('Employee: ');

  userInterface.on("line", async (rawLine) => {
    const line = rawLine.replace('Employee: ', '');
    convo.push({ role: "user", content: line });

    const response = await openaiapi.createChatCompletion({
      model: "gpt-4-turbo",
      messages: convo,
      temperature: 0.9,
      max_tokens: 100,
      top_p: 1,
      frequency_penalty: 0.5,
      presence_penalty: 0.6,
    });

    const answerBot = response.data.choices[0].message.content;
    convo.push({ role: "assistant", content: answerBot });
    console.log("Chat bot: ", answerBot);

    userInterface.prompt();
    userInterface.write('Employee: ');
  });
};

chatLoop();
```

## ДОДАТОК В. Перелік документів на оптичному носії

1. Самойленко\_ДД\_125-20-3\_ПРЕЗ\_КВ\_РОБ.pptx
2. Самойленко\_ДД\_125-20-3\_КВ\_РОБ.docx
3. Самойленко\_ДД\_125-20-3\_КВ\_РОБ.pdf
4. ДОДАТОК Е. Діалоги на основі підказок (prompts).docx
5. Самойленко\_ДД\_125-20-3\_КВ\_РОБ.pdf.p7s

## ДОДАТОК Г. Відгук керівника економічного розділу

Економічний розділ виконаний відповідно до вимог, які ставляться до кваліфікаційних робіт, та заслуговує на оцінку 95б. («відмінно»).

Керівник розділу

\_\_\_\_\_

(підпис)

Дар'я ПІЛОВА

(ім'я, прізвище)

ДОДАТОК Д. Відгук керівника кваліфікаційної роботи

ВІДГУК

на кваліфікаційну роботу студентки групи 125-20-3

Самойленко Діани Дмитрівни

на тему: «Підвищення освіченості персоналу з протидії фішингу з використанням моделі генеративного штучного інтелекту»

Пояснювальна записка складається зі вступу, трьох розділів і висновків, викладених на 94 сторінках.

Метою кваліфікаційної роботи є підвищення обізнаності персоналу щодо протидії фішинговим атак через інтерактивні симуляції з використанням моделі генеративного штучного інтелекту.

Тема кваліфікаційної роботи безпосередньо пов'язана з об'єктом діяльності бакалавра спеціальності 125 «Кібербезпека». Для досягнення поставленої мети в кваліфікаційній роботі вирішуються наступні задачі: аналіз актуальності проблеми соціальної інженерії в контексті кібербезпеки; проведення порівняльного аналізу інструментів генеративного штучного інтелекту; розробка плану тренінгу з протидії фішинговим атакам; створення та реалізація тренувальних шаблонів фішингових атак для різних категорій співробітників; тестування та оцінка ефективності роботи чат-бота.

Розроблено план тренінгу з підвищення освіченості персоналу щодо протидії фішинговим атакам з використанням генеративного штучного інтелекту.

Практичне значення результатів кваліфікаційної роботи полягає у використанні сучасних технологій генеративного штучного інтелекту для автоматизації та створення реалістичних симуляцій фішингових атак, що покращує результати навчання.

Оформлення пояснювальної записки до кваліфікаційної роботи виконано з незначними відхиленнями від стандартів.

За час дипломування Самойленко Д.Д. проявила себе фахівцем, здатним самостійно вирішувати поставлені задачі та заслуговує присвоєння кваліфікації бакалавра за спеціальністю 125 Кібербезпека, освітньо-професійна програма «Кібербезпека».

Рівень запозичень у кваліфікаційній роботі не перевищує вимог «Положення про систему виявлення та запобігання плагіату».

Кваліфікаційна робота заслуговує оцінки 95 (відмінно).

Керівник кваліфікаційної роботи  
Керівник спец. розділу

д.т.н., проф. В.І. Корнієнко  
ст. викл. Д.С. Тимофєєв