

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Сумський державний університет
Факультет електроніки та інформаційних технологій
Кафедра інформаційних технологій

До захисту допущено
В.о. завідувача кафедри

_____ Світлана ВАЩЕНКО

_____ 2024 р.

КВАЛІФІКАЦІЙНА РОБОТА

на здобуття освітнього ступеня бакалавр

зі спеціальності 122 Комп'ютерні науки,

освітньо-професійної програми Інформаційні технології проектування

на тему: Програмний додаток нейромережевої генерації зображень для рекламних кампаній

Здобувача групи ІТ-03-1 Волкова Павла Костянтиновича
(шифр групи) (прізвище, ім'я, по батькові)

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

(підпис)

Павло ВОЛКОВ
(Ім'я та ПРІЗВИЩЕ здобувача)

Керівник доцент кафедри інформаційних технологій, к.т.н., доц., Юлія ПАРФЕНЕНКО

(посада, науковий ступінь, вчене звання, Ім'я та ПРІЗВИЩЕ)

(підпис)

Сумський державний університет
Факультет електроніки та інформаційних технологій
Кафедра інформаційних технологій
Спеціальність 122 Комп'ютерні науки
Освітньо-професійна програма Інформаційні технології проектування

ЗАТВЕРДЖУЮ

В. о. зав. кафедри ІТ

_____ Світлана ВАЩЕНКО

_____ 2024 р.

З А В Д А Н Н Я

НА КВАЛІФІКАЦІЙНУ РОБОТУ БАКАЛАВРА СТУДЕНТУ

Волкова Павла Костянтиновича

1 Тема роботи Програмний додаток нейромережевої генерації зображень для рекламних кампаній

керівник роботи Парфененко Юлія Вікторівна, к.т.н., доцент,

затверджені наказом по університету від 29 04 2024 р. №0588-VI

2 Строк подання студентом роботи 30 05 2024 р.

3 Вхідні дані до роботи технічне завдання, зображення

4 Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити) аналіз предметної області, постановка задачі, аналіз існуючих моделей штучного інтелекту для генерації зображень, моделювання програмного додатку в IDEF0, архітектура додатку, реалізація програмного додатку, адаптація нейронної мережі, тестування роботи додатку

5 Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень) аналіз предметної області, огляд останніх досліджень, аналіз аналогів, модель нейронної мережі, планування робіт, IDEF0 - діаграми, реалізація програмного додатку.

6. Консультанти розділів роботи:

Розділ	Консультант	Підпис, дата	
		Завдання видав	Завдання прийняв

7. Дата видачі завдання _____

КАЛЕНДАРНИЙ ПЛАН

№ п/п	Назва етапів кваліфікаційної роботи	Строк виконання етапів роботи	Примітка
1	Аналіз предметної області	До 09.04.2024	
2	Оформлення планування робіт	До 15.04.2024	
3	Створення технічного завдання	До 20.04.2024	
4	Проектування програмного додатку	До 24.04.2024	
5	Розробка програмного додатку	До 06.05.2024	
6	Тестування програмного додатку	До 20.05.2024	
7	Написання пояснювальної записки	До 25.05.2024	
8	Оформлення презентації	До 29.05.2024	

Студент

(підпис)

Павло ВОЛКОВ

Керівник роботи

(підпис)

к.т.н., доц. Юлія ПАРФЕНЕНКО

АНОТАЦІЯ

Тема кваліфікаційної роботи бакалавра «Програмний додаток нейромережевої генерації зображень для рекламних кампаній».

Пояснювальна записка складається зі вступу, 3 розділів, висновків, списку використаних джерел із 24 найменувань, додатків. Загальний обсяг роботи – 61 сторінка, у тому числі 38 сторінки основного тексту, 2 сторінки списку використаних джерел, 18 сторінок додатків.

У роботі розроблено програмний додаток, який дозволяє генерувати високоякісні зображення за допомогою нейромережевих технологій для використання у рекламних кампаніях. Додаток побудовано на основі моделі Stable Diffusion XL з адаптацією за методом Low-Rank Adaptation (LoRA) та забезпечує зручний інтерфейс для введення текстових запитів, налаштування параметрів генерації та перегляду результатів.

У першому розділі проведено огляд сучасних досліджень у сфері нейромережевої генерації зображень, проаналізовано існуючі моделі та обґрунтовано вибір Stable Diffusion XL з адаптацією методом LoRA для реалізації проекту.

Другий розділ присвячено моделюванню програмного додатку за методологією IDEF0 та побудові UML діаграм.

У третьому розділі описано архітектуру додатку, етапи його розробки, адаптацію нейромережевої моделі та тестування роботи.

Ключові слова: програмний додаток, нейронна мережа, Python, Stable Diffusion, реклама, зображення, LoRA.

ЗМІСТ

Вступ.....	6
1 Аналіз предметної області.....	8
1.1 Огляд останніх досліджень і публікацій.....	8
1.2 Аналіз існуючих моделей штучного інтелекту для генерації зображень.....	11
1.3 Постановка задачі.....	15
1.4 Модель нейронної мережі	18
1.5 Основні параметри генерації зображень.....	20
2 Моделювання та проектування.....	24
2.1 Моделювання програмного додатку в IDEF0	24
2.2 Діаграма варіантів використання програмного додатку	27
3 Розробка програмного додатку	30
3.1 Архітектура додатку.....	30
3.2 Реалізація програмного додатку	31
3.3 Адаптація нейронної мережі.....	33
3.4 Тестування роботи додатку	36
Висновки	39
Список використаних джерел	40
Додаток А. Технічне завдання	43
Додаток Б. Планування робіт.....	50
Додаток В. Лістниг програмного коду.....	58
Додаток Г. Апробація результатів дослідження	61

ВСТУП

Реклама є ключовим елементом успіху будь-якого бізнесу. Вона не тільки інформує потенційних клієнтів про продукти та послуги, але й стимулює інтерес та бажання покупки. З розвитком технологій, зокрема у сфері штучного інтелекту, відкриваються нові можливості для креативності та ефективності в рекламних кампаніях. Одним із таких нововведень є використання нейромереж для генерації зображень, що може кардинально змінити підхід до створення рекламного контенту.

Об'єкт дослідження – процес генерації зображень за допомогою нейронних мереж.

Предмет дослідження охоплює процеси навчання нейромережі за допомогою методу LoRA (Low-Rank Adaptation) на базі моделі Stable Diffusion XL для генерації зображень, які будуть відповідати специфічним вимогам рекламних кампаній. Дослідження зосередиться на аналізі та вдосконаленні алгоритмів навчання, щоб забезпечити високу якість та релевантність генерованих зображень.

Мета роботи полягає у розробці програмного додатку, який за допомогою нейромережевої генерації зображень дозволить створювати унікальні та високоякісні візуальні матеріали для рекламних кампаній. Це дозволить не тільки оптимізувати процес створення рекламного контенту, але й забезпечити його високу адаптивність та персоналізацію.

Для досягнення поставленої мети треба вирішити наступні завдання:

- провести дослідження та аналіз нейромережевих моделей генерації зображень;
- сформулювати технічне завдання та план реалізації програмного продукту;
- за допомогою існуючих методів персоналізувати обрану нейромережу;
- створити програмний додаток генерації зображень, використовуючи персоналізовану модель нейромережі.

Очікувані результати включають створення прототипу програмного додатку, який демонструє можливості нейромережевої генерації зображень для рекламних

кампаній. Структура роботи передбачає детальний огляд літератури, аналіз існуючих рішень, розробку та тестування програмного додатку, а також оцінку його впливу на рекламну індустрію.

Практичне значення роботи полягає у створенні інструменту, який зможе автоматизувати та поліпшити процес створення рекламних матеріалів, зробивши його більш гнучким та економічно вигідним. Такий підхід має потенціал революціонізувати рекламну галузь, пропонуючи новітні рішення для брендів та рекламодавців.

1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ

1.1 Огляд останніх досліджень і публікацій

Нейромережеві технології генерації зображень за останнє десятиліття досягли значного прогресу завдяки появі таких моделей, як Stable Diffusion [1], DALL-E, та Midjourney [2]. Ці потужні моделі здатні створювати високоякісні зображення на основі текстових запитів, відкриваючи нові перспективи для творчості та креативного процесу. Stable Diffusion, зокрема, використовує техніку дифузії латентних просторів, яка ґрунтується на поєднанні варіаційних автокодувальників та дифузійних моделей. На першому етапі застосовується автокодувальник для стиснення вхідного зображення в латентний простір нижчої розмірності. Потім дифузійна модель генерує зображення в цьому латентному просторі, використовуючи текстові умови для керування процесом. Це дозволяє ефективно синтезувати високоякісні складні зображення без необхідності обробки у високовимірному просторі пікселів.

Однак, при всіх своїх перевагах, донавчання цілої моделі, наприклад, Stable Diffusion, є ресурсоємним процесом, що вимагає значних обчислювальних потужностей і великих обсягів даних для тренування. Альтернативним і більш ефективним підходом є метод Low-Rank Adaptation (LoRA), запропонований дослідниками з Microsoft у 2021 році [3]. LoRA дозволяє адаптувати великі моделі, такі як GPT-3 або Stable Diffusion, до конкретних завдань за допомогою низькорангових матриць, що значно зменшує кількість параметрів, які потрібно навчати, і обчислювальні вимоги.

Принцип LoRA полягає в тому, щоб заморозити основні ваги передтренуваної моделі та оптимізувати лише низькорангові матриці розкладання. Це дозволяє значно скоротити кількість тренуваних параметрів, роблячи процес більш ефективним з точки зору обчислень та пам'яті. LoRA дає кілька ключових переваг порівняно з повним перенавчанням моделі: можливість розділяти одну передтренувану модель для різних завдань, підвищення ефективності тренування

завдяки відсутності необхідності розраховувати градієнти для більшості параметрів, економія пам'яті та полегшене переключення між завданнями.

Крім LoRA існують інші перспективні методи адаптації моделей генерації зображень. Одним з них є Textual Inversion [4], який дозволяє отримувати зображення, що відповідають певному поняттю або контексту, шляхом оптимізації вхідного шуму моделі. В статті наголошується, що метод LoRA часто демонструє кращі результати, ніж Textual Inversion, особливо для генерації специфічних об'єктів чи персонажів[5].

Ще одним цікавим підходом є DreamBooth [6], розроблений Google Research, який дозволяє налаштувати модель на невелику кількість зображень для створення персоналізованих зображень, наприклад, з певною людиною або об'єктом. Однак, як зазначено в статті, DreamBooth може бути менш ефективним, ніж LoRA, особливо коли необхідно багато ітерацій тренування чи швидка переналаштовуваність моделі.

Застосування стабільної дифузії та технік точного налаштування, таких як LoRA, для генерації зображень у рекламі та створенні логотипів має кілька ключових переваг:

1. Це спрощує та прискорює процес створення візуального контенту, зменшуючи потребу в дорогих процесах, таких як фотозйомка з залученням моделей та фотографів.
2. Згенеровані зображення можна легко адаптувати під конкретні потреби рекламної кампанії чи бренду завдяки налаштуванню моделі.
3. Підхід забезпечує гнучкість та різноманітність, дозволяючи швидко генерувати безліч варіантів для вибору найкращого.
4. Це скорочує витрати та вимоги до людських ресурсів у порівнянні з традиційними методами створення контенту.

Незважаючи на певні обмеження, застосування технологій генерації зображень на основі тексту, таких як Stable Diffusion, має величезний потенціал для трансформації рекламної галузі та створення цифрового контенту в цілому. Як зазначено в статті [7], використання штучного інтелекту для генерації контенту

(Artificial Intelligence Generated Content, AIGC) швидко набирає популярності та розглядається як один з найбільш перспективних напрямків у сфері машинного навчання та автоматизації.

Однак, разом з технологічним прогресом виникають і нові виклики, які потребують вирішення. Одним з ключових питань є відсутність чітких правових норм та законодавчого регулювання для контенту, згенерованого за допомогою штучного інтелекту. Питання авторських прав на такий контент залишаються невизначеними в багатьох країнах та регіонах, що може призвести до правових суперечок у майбутньому.

Крім того, існують сумніви щодо достовірності та автентичності зображень, згенерованих штучним інтелектом. У певних сферах, таких як мода чи реклама, використання згенерованих зображень замість реальних може викликати скептицизм у споживачів щодо справжності представленого продукту. Це може зашкодити основній меті реклами - допомогти покупцям візуалізувати реальний вплив товару.

Незважаючи на ці проблеми, переваги технологій AIGC є безперечними. При належному використанні вони можуть спричинити революцію в рекламній індустрії, значно знизивши витрати та потребу в людських ресурсах. Крім того, AIGC забезпечує більш ефективний та гнучкий підхід до створення привабливого візуального контенту, який тісно пов'язаний з ідентичністю бренду та його повідомленням.

Очікується, що з подальшим розвитком технологій AIGC питання, пов'язані з їх застосуванням, будуть вирішені, а більш жорсткі правові рамки будуть встановлені. За умови обережного та етичного використання ці технології, безсумнівно, здатні принести значні економічні вигоди в майбутньому, революціонізувавши різні галузі. Це робить їх перспективним інструментом у сфері створення цифрового контенту.

Таким чином створення програмного додатку, який інтегрує методи нейромережевої генерації зображень для створення візуального контенту для

рекламних кампаній є актуальною задачею, що може забезпечити ефективне та економічне вирішення потреб рекламної індустрії.

1.2 Аналіз існуючих моделей штучного інтелекту для генерації зображень

На ринку наявні кілька потужних моделей для генерації зображень на основі текстових запитів, серед яких DALL-E 3 [8], Midjourney та Stable Diffusion. Розглянемо їх переваги та недоліки в контексті створення візуального контенту для рекламних кампаній.

DALL-E 3, розроблена компанією OpenAI, є визначною моделлю штучного інтелекту для перетворення тексту на зображення завдяки покращеному розумінню складних текстових підказок. Порівняно з попередником DALL-E 2, DALL-E 3 демонструє вищу точність і якість у генерації зображень за однакових підказок. Унікальною особливістю DALL-E 3 є інтеграція з розмовними можливостями ChatGPT, що створює середовище взаємодії, в якому користувачі можуть уточнювати підказки за допомогою діалогу для більш інтуїтивного процесу створення зображень.

Стосовно доступності, DALL-E 3 наразі доступна для клієнтів ChatGPT Plus і Enterprise, причому користувачі мають повні авторські права на створені зображення. Це розширення прав є надзвичайно важливим, оскільки дозволяє приватним особам і компаніям вільно використовувати ці зображення для особистих проєктів, комерційного використання чи подальших творчих починань.

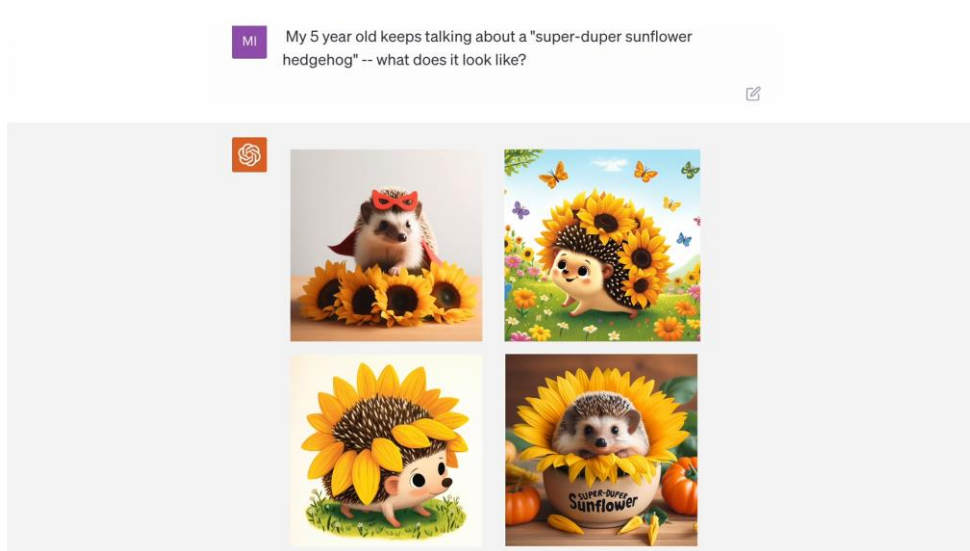


Рисунок 1.1 – Приклад роботи DALL-E 3 в ChatGPT Plus

Також DALL-E 3 доступна з використанням API через HTTP запити з будь-яких мов програмування, через офіційні Python та Node.js бібліотеки, або бібліотеки що підтримує спільнота[9].

Приклад використання офіційної бібліотеки на мові Python[10]:

```
from openai import OpenAI
client = OpenAI()
response = client.images.generate(
    model="dall-e-3",
    prompt="a white siamese cat",
    size="1024x1024",
    quality="standard",
    n=1,
)
image_url = response.data[0].url
```

Midjourney – це інноваційна програма, розроблена компанією Midjourney, Inc. для створення візуального контенту безпосередньо з текстових описів. Подібно до DALL-E та Stable Diffusion, Midjourney використовує потужність природної

мови для формування та втілення візуальних ідей. Сервіс є доступним, використовуючи Discord як свій інтерфейс, що дозволяє користувачам майже миттєво створювати зображення з текстових підказок без необхідності встановлювати додаткове програмне забезпечення.

Midjourney пропонує чотири рівні підписки з гнучкими варіантами платежів, забезпечуючи доступ до різних функцій та комерційних умов використання. Універсальність Midjourney полягає в його здатності генерувати широкий спектр стилів, від гіперреалістичних до абстрактних, роблячи його потужним інструментом для творчих професіоналів, включаючи художників, дизайнерів і маркетологів у рекламній сфері.

Midjourney не має програмного доступу через API [2], використовувати дану модель можливо тільки через бота в Discord, або через web версію, яка доступна лише користувачам, які згенерували понад 1000 зображень.

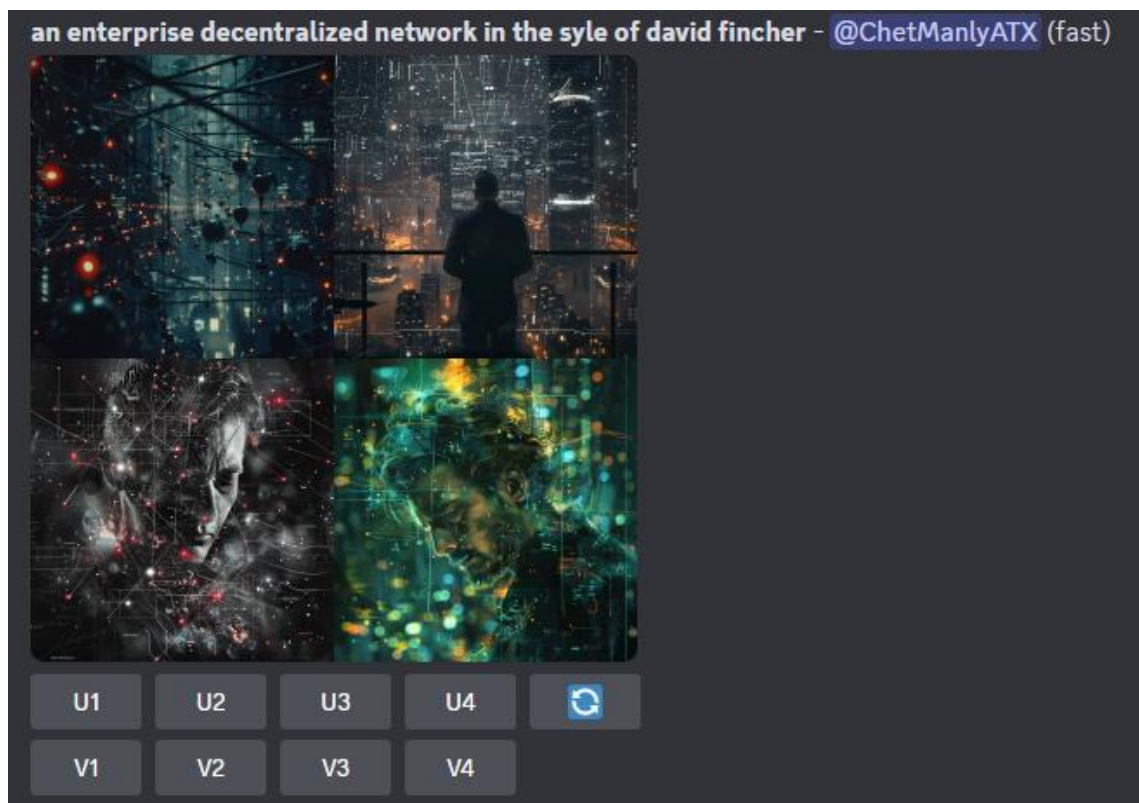


Рисунок 1.2 – Приклад використання Midjourney в Discord

Stable Diffusion є знаковою розробкою в галузі штучного інтелекту, що втілює потужну модель дифузії тексту в зображення. Ця модель вирізняється здатністю генерувати не лише зображення високої якості, але й вражаюче фотореалістичні результати. Вона покликана демократизувати процес створення мистецтва, пропонуючи засоби для створення захоплюючих візуальних ефектів з текстових підказок для широкої аудиторії з безпрецедентною швидкістю.

Поява Stable Diffusion XL стала значним кроком вперед в еволюції моделі. Ця вдосконалена версія спрощує процес створення складних зображень, оскільки вимагає менш детальних підказок для створення конкретних та описових візуальних образів. Унікальним аспектом Stable Diffusion XL є її здатність інтегрувати та генерувати текст всередині самих зображень, що розширює можливості створення зображень та історій, які вони можуть розповісти. Покращення композиції зображень і генерації людських облич сприяють створенню результатів, які вражають не лише своєю реалістичністю, але й художньою якістю.

Випускаючи ці моделі з відкритим вихідним кодом, Stability AI гарантує, що творці, розробники та дослідники матимуть свободу будувати, модифікувати та інтегрувати модель у різноманітні додатки.

Stable Diffusion XL доступна для завантаження на платформі Hugging Face[11] у форматі safetensors який представляє собою збережені тензори неймережі у стиснутому вигляді. Один із способів використання моделі це запуск за допомогою Python бібліотек – PyTorch[12] та diffusers, приклад програмного коду[11]:

```
from diffusers import DiffusionPipeline
import torch
pipe =
DiffusionPipeline.from_pretrained("stabilityai/stable-diffusion-
xl-base-1.0", torch_dtype=torch.float16, use_safetensors=True,
variant="fp16")
```

```

pipe.to("cuda")
prompt = "An astronaut riding a green horse"
images = pipe(prompt=prompt).images[0]

```

Таблиця 1.1 – Порівняльна характеристика аналогічних моделей

Критерій порівняння	DALL-E 3	Midjourney	Stable Diffusion
Якість генерованих зображень	Висока	Висока	Висока
Відкритий вихідний код	Ні	Ні	Так
Можливість персоналізації	Ні	Ні	Так (за допомогою LoRA, Textual Inversion, DreamBooth)
Інтеграція у власні додатки	Обмежена	Відсутня	Можлива
Вартість використання	Комерційна	Комерційна	Безкоштовна (за відкритим вихідним кодом)
Розробник	OpenAI	Midjourney	Stability AI, Runway

Таким чином, хоча DALL-E 3 та Midjourney забезпечують високу якість генерованих зображень, лише Stable Diffusion надає можливість персоналізації результатів генерації за допомогою методів, описаних у розділі 1.1. Це робить Stable Diffusion найбільш привабливим вибором для створення програмного додатку для генерації візуального контенту для рекламних кампаній, оскільки дозволяє адаптувати модель до специфічних потреб кожної окремої кампанії

1.3 Постановка задачі

Метою проекту є створення програмного додатку для генерації зображень за допомогою нейромережових технологій, призначеного для використання у рекламних кампаніях. Додаток має надавати зручний інтерфейс для введення

текстових запитів та налаштування параметрів генерації зображень на основі моделей генерації зображень, таких як Stable Diffusion. Це дозволить рекламним агентствам, маркетологам та дизайнерам швидко та ефективно створювати високоякісний візуальний контент для своїх рекламних кампаній.

Для реалізації даного проекту будуть використані наступні засоби та технології:

- Python: Основна мова програмування для реалізації логіки додатку, включаючи обробку текстових запитів та взаємодію з нейромережею.
- Gradio: бібліотека для створення веб-інтерфейсів, який дозволяє швидко розробляти та тестувати зручні інтерфейси користувача для роботи з нейромережами.
- stable-diffusion-webui від AUTOMATIC1111: Інтерфейс для взаємодії з моделлю Stable Diffusion, що забезпечує зручний спосіб інтеграції та налаштування нейромережевої моделі.
- kohya_ss: Набір інструментів для тренування нейромережі, який буде використаний для адаптації моделі Stable Diffusion відповідно до специфічних вимог проекту.

Для досягнення поставленої мети необхідно вирішити наступні завдання:

- Розробити технічне завдання на проект, в якому докладно описати функціональні та системні вимоги до програмного додатку, а також етапи розробки. Технічне завдання має містити детальний опис вимог до структури, функціоналу, зручності використання, а також інформацію про необхідне апаратне та програмне забезпечення для коректної роботи додатку.
- Виконати планування робіт, визначити необхідні ресурси та проаналізувати можливі ризики для успішної реалізації проекту. Планування включає створення діаграм Ганта, визначення структури робіт (WBS), організаційної структури (OBS), а також аналіз та розробку стратегій мінімізації ризиків.

- На основі проведеного дослідження виконати адаптацію нейромережевої моделі Stable Diffusion XL за допомогою методу Low-Rank Adaptation.

- Розробити та інтегрувати функціонал для введення текстових запитів, налаштування параметрів генерації, перегляду результатів та збереження згенерованих зображень. Забезпечити зручний інтерфейс для введення текстових запитів, вибору моделі та налаштування параметрів генерації, таких як розмір зображення, кількість кроків та інші відповідні параметри.

- Створити зручний та інтуїтивно зрозумілий графічний інтерфейс користувача, що відповідає сучасним стандартам UI/UX дизайну. Інтерфейс повинен бути зручним у використанні, забезпечувати швидкий доступ до основних функцій, а також надавати можливість зручного перегляду та збереження згенерованих зображень.

- Провести тестування програмного додатку, усунути виявлені недоліки та забезпечити стабільну роботу. Ретельне тестування програмного продукту є критично важливим для гарантії його коректної роботи та задоволення потреб користувачів.

- Підготувати необхідну документацію, яка включає інструкції з використання, опис функцій та можливостей програмного додатку, а також технічну документацію для подальшої підтримки та розвитку продукту.

Більш детальні вимоги до проекту, включаючи функціональні та системні вимоги, потреби користувачів та інші аспекти, описані у технічному завданні на розробку програмного додатку (Додаток А).

Реалізація даного проекту дозволить створити потужний інструмент для автоматизованої генерації високоякісних зображень за допомогою передових нейромережевих технологій, що значно спростить процес створення візуального контенту для рекламних кампаній та підвищить ефективність роботи відповідних фахівців.

1.4 Модель нейронної мережі

Модель нейронної мережі Stable Diffusion, також відома як Latent Diffusion Model (LDM), представляє сучасний підхід до генерації зображень з високою роздільною здатністю, використовуючи методи латентного простору для зменшення обчислювальних витрат при збереженні високої якості результатів.

Основна ідея Latent Diffusion Model полягає в тому, щоб виконувати процес дифузії не в оригінальному просторі пікселів, а в латентному просторі, який є нижчої розмірності та зберігає основні семантичні характеристики зображень. Цей підхід дозволяє зменшити обчислювальну складність як при тренуванні моделі, так і під час генерації зображень.

LDM складається з двох основних етапів:

Автокодування: спочатку тренується автоенкодер, який перетворює зображення в компактне латентне представлення. Автоенкодер складається з енодера, який зменшує розмірність даних, і декодера, який відновлює зображення з латентного простору.

Важливою особливістю є те, що автоенкодер навчається один раз і може бути використаний для різних задач без необхідності повторного навчання [1].

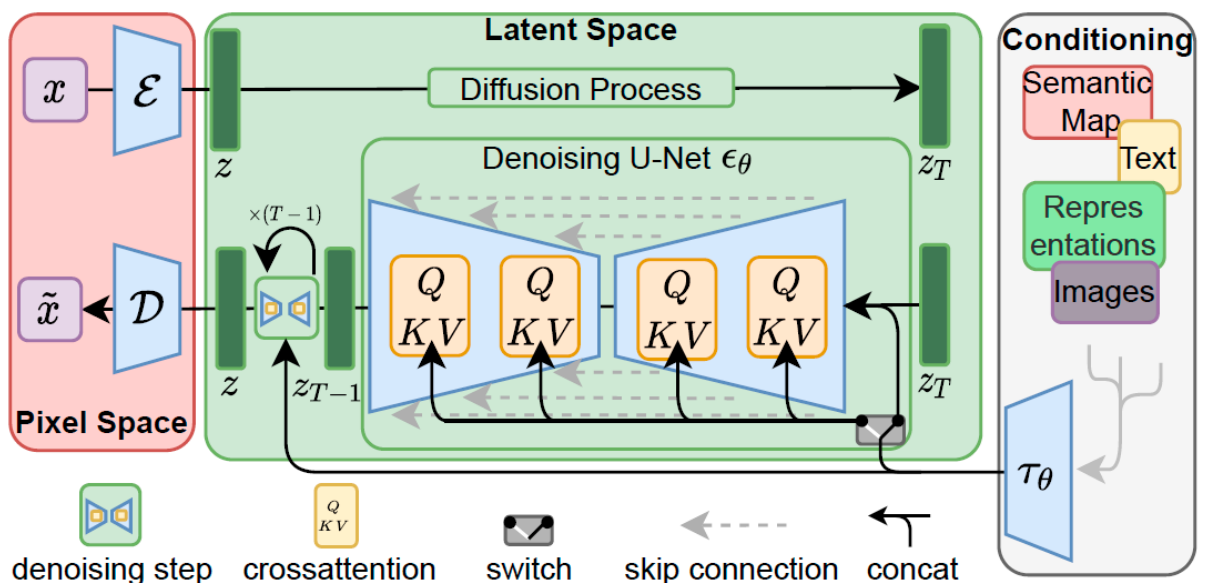


Рисунок 2.4 – Архітектура LDM [1]

Дифузійна модель у латентному просторі: на другому етапі навчається власне дифузійна модель, але вже в латентному просторі, що значно зменшує обчислювальні витрати. Дифузійна модель здійснює процес генерації шляхом поступового внесення та видалення шуму в латентному просторі, використовуючи U-Net архітектуру для реконструкції зображень.

Переваги підходу

Ефективність: завдяки роботі в латентному просторі, LDM потребує менше обчислювальних ресурсів у порівнянні з традиційними дифузійними моделями, що працюють в оригінальному просторі пікселів.

Висока якість: модель забезпечує високоякісні та деталізовані зображення, оскільки процес компресії зберігає основні семантичні деталі.

Гнучкість: одноразове навчання автоенкодера дозволяє повторно використовувати його для різних задач, таких як умовна генерація зображень, “суперрозольція”(super-resolution), “інпейнтинг”(inpainting) та інші.

Скорочення обчислювальних витрат: модель значно знижує вимоги до обчислювальних ресурсів під час генерації зображень, що робить її доступною для ширшого кола дослідників та практиків[1].

Stable Diffusion XL (SDXL) є розширеною версією моделі Stable Diffusion, яка орієнтована на досягнення ще вищої якості зображень та більшої деталізації. SDXL використовує глибшу архітектуру та більшу кількість параметрів, що дозволяє їй краще захоплювати складні структури та текстури в зображеннях.

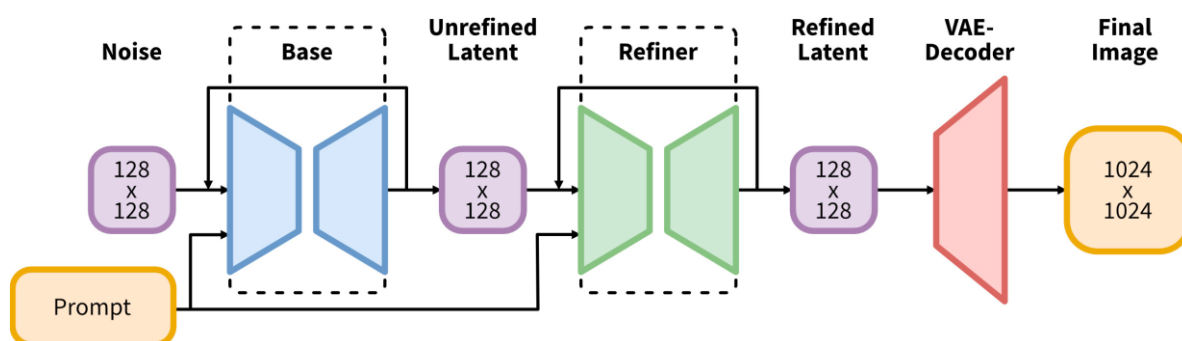


Рисунок 2.5 – Візуалізація двоступеневого процесу генерації [19]

Основні особливості SDXL

Покращена архітектура: SDXL включає більше шарів та більшу кількість нейронів у кожному шарі, що сприяє більш точному захопленню деталей та складних структур в зображеннях.

Висока роздільна здатність: модель здатна генерувати зображення з дуже високою роздільною здатністю, що є важливим для додатків, де важлива якість та чіткість деталей.

Гнучкість у застосуванні: SDXL може бути адаптована для різних задач, включаючи генерацію фотореалістичних зображень, художніх творів, та інших типів візуального контенту.

Використання у проекті

У даній роботі використовуватиметься модель SDXL як основна модель для генерації зображень. Важливо зазначити, що в цьому випадку буде використано лише основну моделі без додаткового кроку з *refiner model*[19]. Це дозволить спростити процес генерації та зменшити обчислювальні витрати, зберігаючи при цьому високу якість результатів.

1.5 Основні параметри генерації зображень

При генерації зображень за допомогою моделей Stable Diffusion важливу роль відіграють різні параметри генерації. Два ключові параметри, які суттєво впливають на якість і характеристику згенерованих зображень, це CFG Scale та Sampling Steps.

CFG Scale (Classifier-Free Guidance Scale) є важливим параметром, який визначає ступінь впливу умовних даних на процес генерації зображень. Цей параметр використовується в методі *classifier-free guidance*, що дозволяє керувати генерацією зображень без необхідності використовувати окремий класифікатор.



Рисунок 2.6 – Classifier-Free Guidance Scale[25]

На рис. 2.6 зображено ефект впливу наведення на суміш трьох гаусівських величин, де кожен компонент суміші представляє дані, що відповідають певному класу. Крайній лівий графік – некерована гранична величина. Зліва направо – величини сумішей нормалізованих керованих умовних значень зі збільшенням сили наведення(CFG Scale).

CFG Scale регулює баланс між умовними (керованими) і безумовними (некерованими) генераціями. Вищі значення CFG Scale означають більший вплив умовних даних на результати генерації, тоді як нижчі значення надають більше свободи моделі, що може призводити до більш різноманітних, але менш керованих результатів.

Типові значення CFG Scale коливаються в діапазоні від 5 до 15. Вибір конкретного значення залежить від специфічних вимог до якості та керованості зображень. Наприклад, для більш контрольованих та чітких результатів може бути використане значення близьке до 8 або вище.

Вищі значення CFG Scale можуть покращити деталізацію та структуру зображень, але надмірно високі значення можуть призвести до втрати різноманітності та поява артефактів через надмірне пристосування до умовних даних.

Sampling Steps є іншим критично важливим параметром, який визначає кількість ітерацій, що використовуються для генерації зображення в процесі дифузії. Кількість кроків впливає на якість та час, необхідний для генерації зображень.

Даний параметр визначає кількість ітерацій дифузійного процесу, через які проходить модель, щоб згенерувати кінцеве зображення. Більша кількість кроків зазвичай призводить до кращої якості зображення, оскільки модель має більше можливостей для точного відновлення деталей зображення.

Зазвичай значення Sampling Steps варіюються від 20 до 400 кроків. Вибір конкретної кількості залежить від необхідного балансу між якістю зображення та обчислювальними витратами. Для більшості задач достатньо 25-70 кроків.

Кількість ітерацій прямо пропорційно впливає на час, необхідний для генерації зображення. Порівняння комбінацій параметрів, які впливають на результат генерації зображень, показано на рис. 2.7.

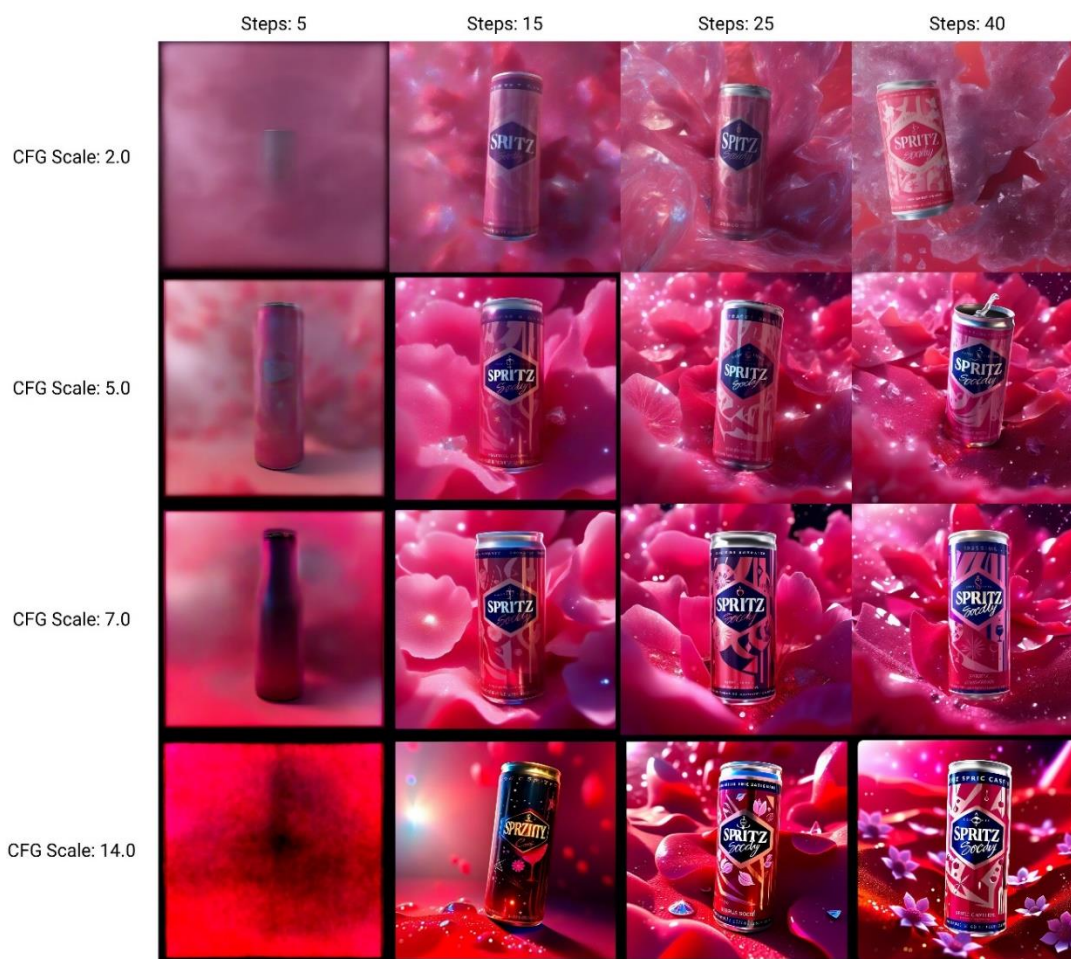


Рисунок 2.7 – Порівняння комбінацій параметрів

Таким чином, параметри CFG Scale та Sampling Steps є ключовими елементами управління процесом генерації зображень в моделях Stable Diffusion.

Їх правильне налаштування дозволяє досягти оптимального балансу між якістю, деталізацією та обчислювальними витратами, що є важливим для успішного використання цих моделей у різних практичних застосуваннях.

2 МОДЕЛЮВАННЯ ТА ПРОЕКТУВАННЯ

2.1 Моделювання програмного додатку в IDEF0

Моделювання програмного додатку для генерації зображень за допомогою нейромережових технологій виконується за допомогою методології IDEF0. Дана методологія дозволяє описати функціональні аспекти системи, її структуру та взаємодію компонентів.

На рис. 2.1 представлена контекстна діаграма процесу генерації зображень в нотатції IDEF0, яка моделює основні компоненти та процеси програмного додатку для генерації зображень.

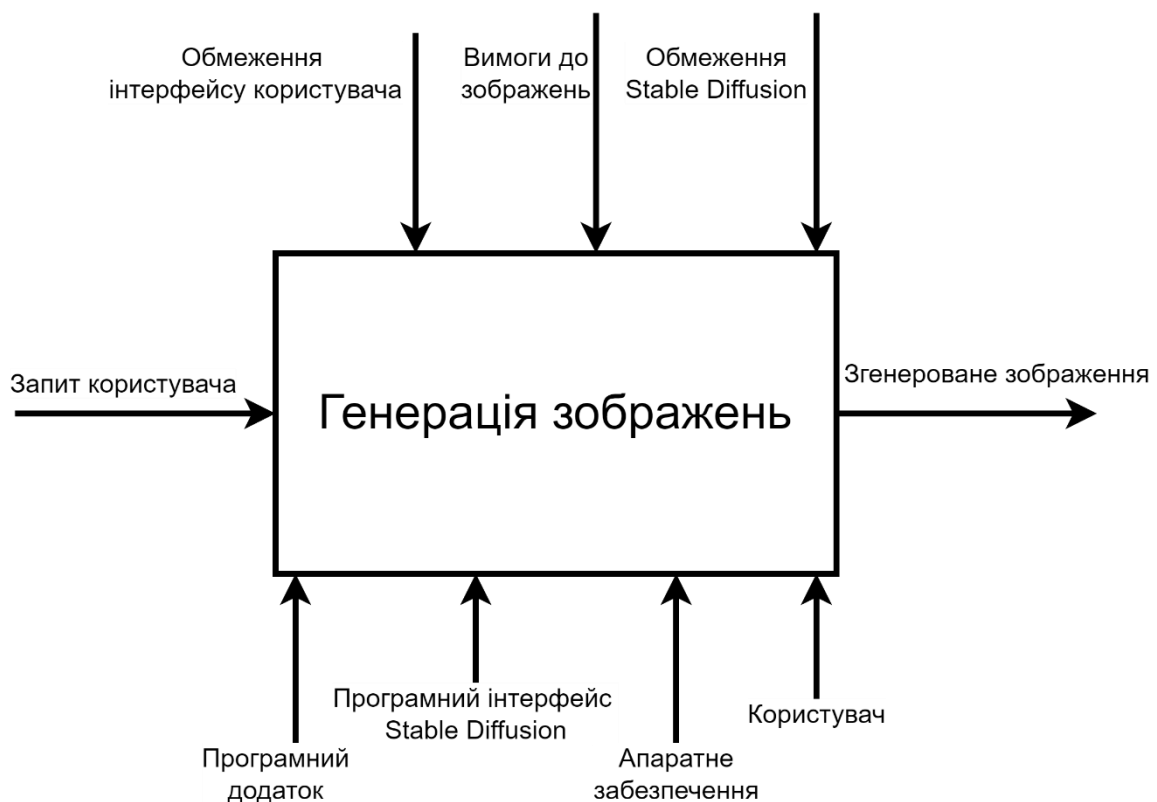


Рисунок 2.1 – Контекстна діаграма процесу генерації зображень

Діаграма складається з наступних блоків:

1. Запит користувача: Це вхідний компонент, який представляє текстовий опис зображення, яке хоче отримати користувач.

2. Вимоги до зображень: Це обмеження та специфікації, які накладаються на процес генерації зображень.
3. Обмеження Stable Diffusion: Це обмеження, які визначають можливості та обмеження нейромережевої моделі Stable Diffusion.
4. Генерація зображень: Основний процес, який відповідає за створення зображень на основі введеного користувачем текстового опису та визначених вимог. Процес включає використання нейромережевої моделі Stable Diffusion та програмного інтерфейсу для взаємодії з нею.
5. Згенероване зображення: Вихідний компонент, який представляє зображення, створене на основі запиту користувача. Це зображення відповідає всім зазначеним вимогам та обмеженням.
6. Користувач: Це користувач, який взаємодіє з програмним додатком. Користувач вводить текстовий запит, переглядає згенеровані зображення та може зберігати їх для подальшого використання.
7. Програмний додаток: Це програмний засіб, який реалізує всі функції, необхідні для генерації зображень. Він включає в себе інтерфейс користувача, API для взаємодії з нейромережею та інші компоненти, що забезпечують коректну роботу додатку.
8. Програмний інтерфейс Stable Diffusion: Це інтерфейс, який дозволяє програмному додатку взаємодіяти з нейромережею Stable Diffusion. Він надає функції для введення параметрів генерації, запуску процесу генерації та отримання результатів.
9. Апаратне забезпечення: Це обчислювальні ресурси, необхідні для роботи нейромережевої моделі та програмного додатку. Включає в себе графічний процесор (GPU) та інші компоненти, необхідні для швидкої та ефективної генерації зображень.

Було проведено декомпозицію контекстної діаграми на чотири блоки.

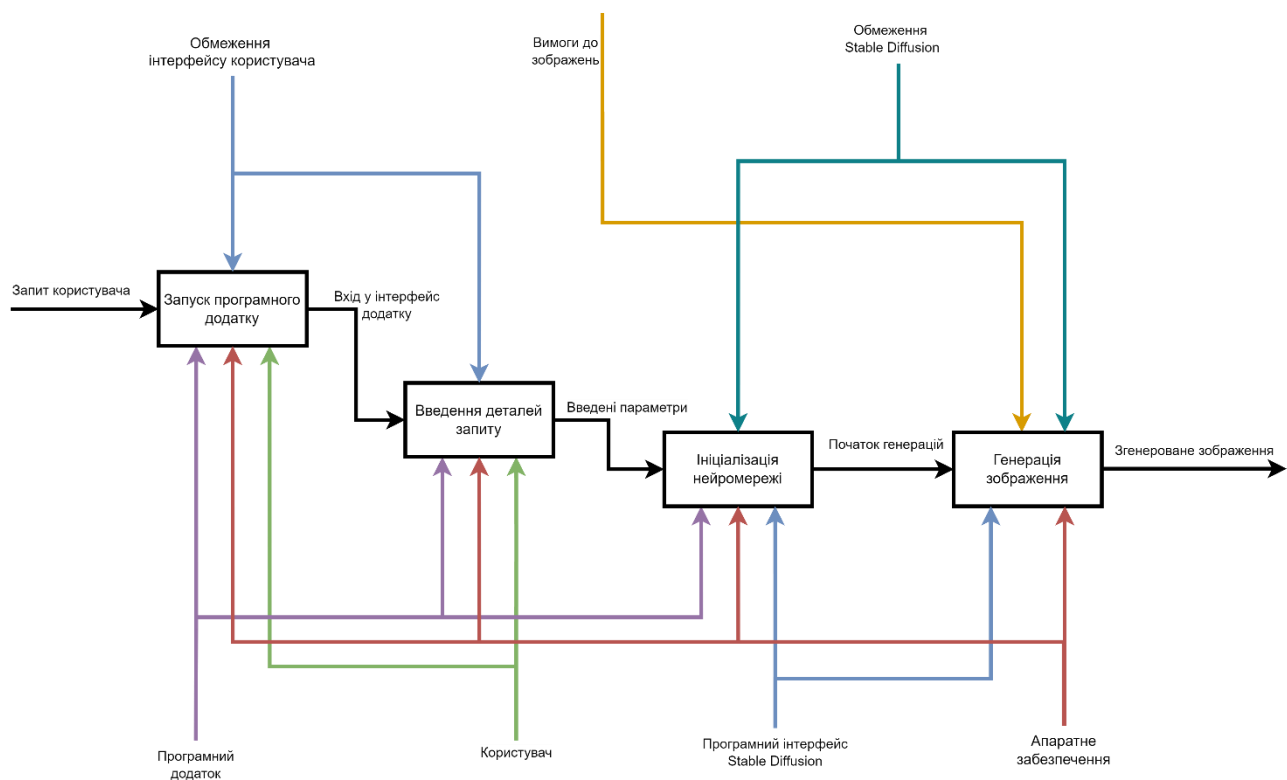


Рисунок 2.2 – Діаграма декомпозиції першого рівня

Діаграма декомпозиції першого рівня процесу "Генерація зображень" в нотації IDEF0 представляє основні кроки, необхідні для створення зображення за допомогою нейромережі. Ця діаграма відображає взаємодію між користувачем, програмним додатком, інтерфейсом користувача, нейромережею та апаратним забезпеченням. Ось опис кожного елемента та взаємодії на діаграмі:

Запуск програмного додатку: користувач запускає програмний додаток, який відповідає за генерацію зображень. На цьому етапі застосовуються обмеження інтерфейсу користувача.

Введення деталей запиту: користувач вводить деталі запиту через інтерфейс додатку. Заповнені параметри запиту передаються далі в процес.

Ініціалізація нейромережі: програмний додаток ініціалізує нейромережу.

Генерація зображення: нейромережа генерує зображення відповідно до параметрів запиту. На цьому етапі враховуються вимоги до зображень та обмеження Stable Diffusion. Генерація зображення також залежить від апаратного забезпечення.

Згенероване зображення: процес завершується отриманням згенерованого зображення, яке передається користувачу.

Взаємодії та обмеження:

1. Обмеження інтерфейсу користувача: впливають на початковий етап запуску програмного додатку.
2. Вимоги до зображень: враховуються під час генерації зображення.
3. Обмеження Stable Diffusion: впливають на процес ініціалізації нейромережі та генерації зображення.
4. Апаратне забезпечення: забезпечує необхідні ресурси для процесу генерації зображення.

Ця діаграма показує, як різні компоненти взаємодіють між собою для досягнення кінцевої мети - генерації зображення відповідно до запиту користувача.

2.2 Діаграма варіантів використання програмного додатку

Діаграма варіантів використання показує основні сценарії взаємодії користувача з програмним додатком для нейромережевої генерації зображень. Вона включає основні варіанти використання, які користувач може виконувати за допомогою додатку.

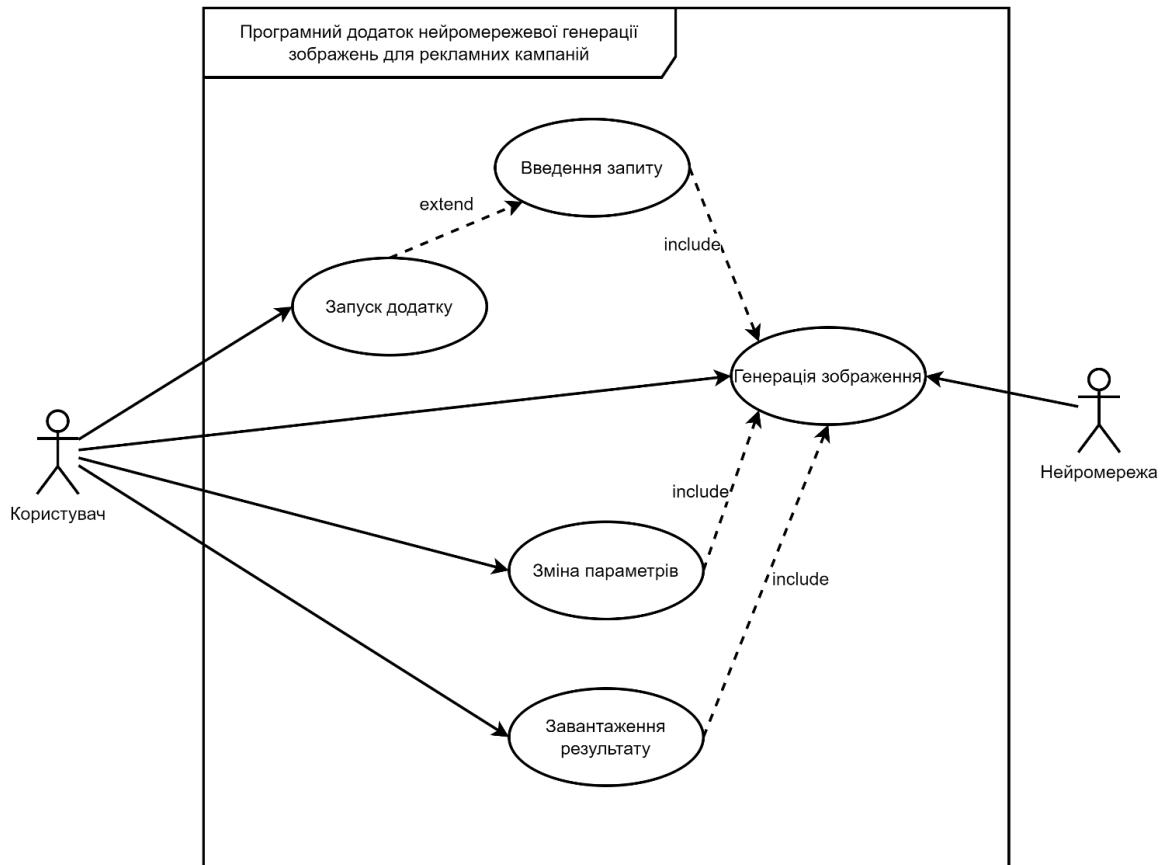


Рисунок 2.3 – Діаграма варіантів використання програмного додатку

На діаграмі варіантів використання (рис. 2.3) зображені основні дії, які можуть виконувати користувачі додатку.

Таблиця 2.1 – Ключові актори

№	Актор	Опис
1	Користувач	Користувач, який взаємодіє з програмним додатком, вводить запити, налаштовує параметри генерації та зберігає результати.
2	Неймережа	Неймережева модель, яка здійснює процес генерації зображень на основі текстових запитів та параметрів.

Таблиця 2.2 – Варіанти використання

№	Назва	Опис
1	Запуск додатку	Користувач запускає програмний додаток для генерації зображень.
2	Введення запиту	Користувач вводить текстовий запит, який описує бажане зображення.
3	Зміна параметрів	Користувач налаштовує параметри генерації, такі як розмір зображення та інші.
4	Генерація зображення	Додаток генерує зображення на основі введеного текстового запиту та налаштованих параметрів.
5	Завантаження результату	Користувач завантажує та зберігає згенероване зображення на свій пристрій.

Таким чином, діаграма варіантів використання та відповідні описи варіантів використання дають чітке розуміння функціональних можливостей програмного додатку та взаємодії користувача з ним.

3 РОЗРОБКА ПРОГРАМНОГО ДОДАТКУ

3.1 Архітектура додатку

Архітектура розробленого програмного додатку нейромережевої генерації зображень для рекламних кампаній представлена на рис. 3.1.

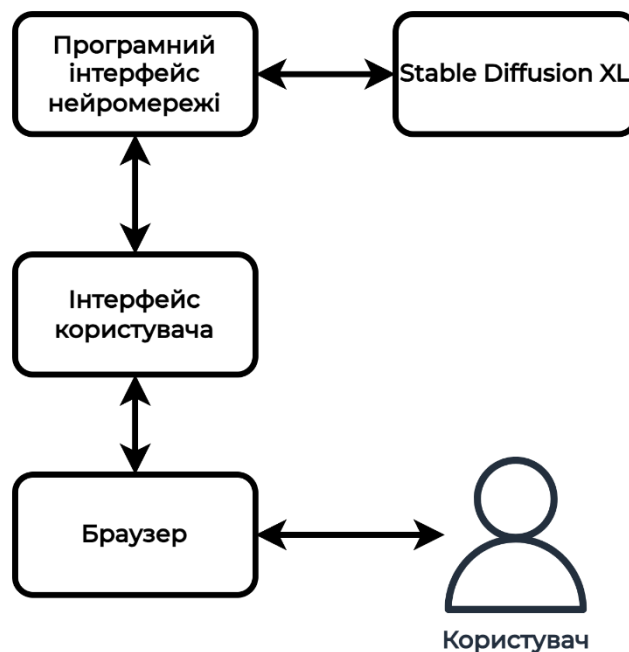


Рисунок 3.1 – Архітектура додатку

Інтерфейс користувача (ІК) є точкою доступу до додатку для користувачів. Він дозволяє користувачам вводити текстовий запит, який описує бажане зображення, а також вибрати параметри генерації. ІК також відображає згенеровані зображення та дозволяє користувачам зберігати їх.

Неймережа є основним компонентом додатку, який відповідає за генерацію зображень. У цьому додатку використовується неймережа Stable Diffusion XL. Підключення до неймережі здійснюється через API, що дозволяє інтегрувати модель безпосередньо у додаток.

Програмний інтерфейс нейромережі (API) – це набір визначень та функцій, які дозволяють програмному забезпеченню взаємодіяти з нейромережею. Він надає абстрактний шар між нейромережею та розробником додатку, приховуючи складність реалізації нейромережі та дозволяючи зосередитися на використанні її функціональних можливостей. У даній роботі використовується `stable-diffusion-webui` від AUTOMATIC1111, який забезпечує зручний інтерфейс для взаємодії з моделлю Stable Diffusion XL. Він надає розробникам інструменти для налаштування параметрів генерації, вибору моделей, а також отримання зворотного зв'язку про якість згенерованих зображень. Використання `stable-diffusion-webui` спрощує процес інтеграції та налаштування нейромережі, дозволяючи розробникам швидко і ефективно впроваджувати нові функції та покращення.

Браузер дозволяє користувачам отримати доступ до додатку, відображаючи інтерфейс користувача.

Таким чином, архітектура додатку включає кілька основних компонентів, які працюють разом для забезпечення ефективної генерації зображень на основі текстових описів.

3.2 Реалізація програмного додатку

Реалізація програмного додатку нейромережевої генерації зображень для рекламних кампаній передбачає використання мови програмування Python і бібліотеки Gradio для створення інтерфейсу користувача, а також репозиторію `stable-diffusion-webui` від AUTOMATIC1111 для взаємодії з нейромережею Stable Diffusion XL.

Основні компоненти реалізації описані нижче.

Інтерфейс користувача.

Використання бібліотеки Gradio для створення веб-інтерфейсу, що забезпечує простоту та інтерактивність у користуванні. Gradio дозволяє швидко

створювати веб-додатки, які взаємодіють з моделями машинного навчання, тому підходить для поставленої задачі.

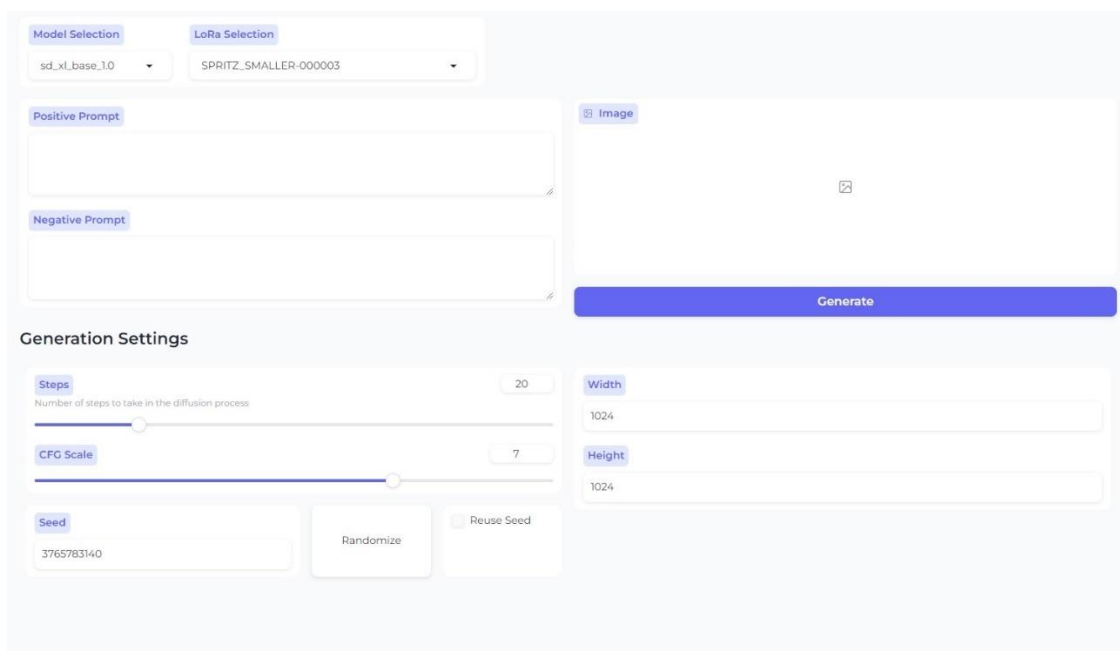


Рисунок 3.1 – Інтерфейс програмного додатку

Інтерфейс дозволяє користувачам вводити текстовий запит для генерації зображень та вибирати параметри, кількість кроків генерації, розмір зображення та інші специфічні налаштування.

Взаємодія з нейромережею.

Використання `stable-diffusion-webui`, який надає зручний спосіб взаємодії для роботи з моделями Stable Diffusion. Цей інструмент забезпечує доступ до функціоналу нейромережі, дозволяючи здійснювати генерацію зображень на основі текстових запитів.

Stable Diffusion XL – це потужна модель для генерації зображень, що використовує глибокі нейронні мережі для створення високоякісних зображень з текстових описів. Інтеграція з `stable-diffusion-webui` спрощує цей процес, надаючи готові інструменти для роботи з моделлю.

Обробка запитів та генерація зображень.

Зв'язок між Gradio та `stable-diffusion-webui` реалізовано через API виклики.

Приклад такого виклику:


```
sd_parameters = {
    "prompt": prompt,
    "negative_prompt": prompt_negative,
    "seed": seed_value,
    "sampler_name": "Euler a",
    "steps": steps,
    "cfg_scale": cfg_scale,
    "width": width,
    "height": height,
}

response =
requests.post("http://127.0.0.1:7860/sdapi/v1/txt2img", json=sd_parameters)
```

Це дозволяє передавати текстові запити від користувачів до нейромережі та отримувати згенеровані зображення у відповідь.

Вхідні дані (текстові запити) передаються до нейромережі, де обробляються моделлю Stable Diffusion XL для генерації відповідних зображень.

3.3 Адаптація нейронної мережі

У цьому підрозділі буде розглянуто процес створення LoRA (Low-Rank Adaptation) для моделі Stable Diffusion XL, використовуючи інструменти `ko_hu_a_ss` [24]. LoRA дозволяє адаптувати великі нейронні мережі для специфічних задач з меншою кількістю параметрів, що значно зменшує обчислювальні витрати та підвищує ефективність тренування.

LoRA (Low-Rank Adaptation) – це метод, який дозволяє адаптувати великі попередньо натреновані моделі до нових задач, зменшуючи кількість необхідних параметрів. Цей підхід використовується для зниження обчислювальної складності та збереження високої якості результатів. Основна ідея LoRA полягає в тому, щоб

представити адаптаційні ваги у вигляді низькорівневої матриці, що додається до існуючих ваг моделі.

Перед початком роботи необхідно встановити всі необхідні інструменти та налаштувати середовище. Основним інструментом для створення LoRA для Stable Diffusion XL є `kohya_ss`. Для встановлення `kohya_ss` необхідно мати Python та декілька інших бібліотек.

Щоб налаштувати проект для Windows, потрібно виконати наступні кроки:

1. Відкрити термінал і перейти до потрібного каталогу встановлення.
2. Клонувати сховище, виконавши наступну команду:

```
git clone --recursive https://github.com/bmaltais/kohya_ss.git
```

5. Перейти до каталогу `kohya_ss`:

```
cd kohya_ss
```

6. Запустити один з наведених нижче скриптів налаштування, виконавши наступну команду:

Для систем, де встановлено лише python 3.10.11:

```
.\setup.bat
```

Для систем, в яких встановлено декілька випусків python:

```
.\setup-3.10.bat
```

На кроці прискорення конфігурації використовуємо значення за замовчуванням. Кількість VRAM на графічному процесорі не впливає на значення, що використовуються.

Підготовка даних

Перед тренуванням моделі необхідно підготувати дані. Для цього слід зібрати набір зображень, що відповідають завданням рекламної кампанії. Зображення повинні бути анотовані, що дозволить моделі краще зрозуміти, які саме елементи є важливими для конкретних задач.

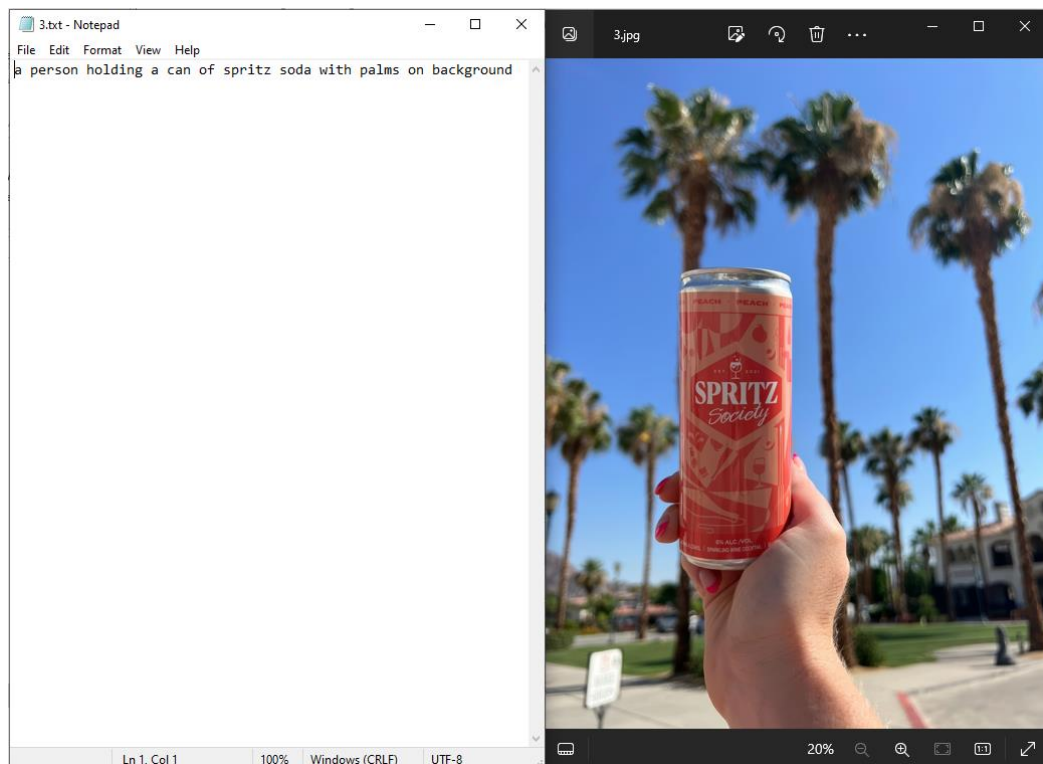


Рисунок 3.2 – Приклад анотованого зображення

Тренування LoRA

Процес тренування LoRA моделі для Stable Diffusion XL включає декілька кроків:

1. Налаштування параметрів тренування.
2. Запуск тренування.
3. Моніторинг процесу.

Основні параметри, за якими проведено тренування:

Repeats: 5;

Train batch size: 1;

Epoch: 5;

LR Scheduler: constant;

Learning rate: 0.0001;

LR warmup: 0;

Network Rank (Dimension): 256;

Результати адаптації

Після завершення тренування важливо оцінити якість отриманих моделей, котрі були збережені на кожній епохі тренування, порівняння результатів зображено на рис. 3.3.



Рисунок 3.3 – Порівняння кожної епохи тренування

Для інтеграції в програмний додаток було обрано модель третьої епохи, оскільки зображення згенеровані за допомогою неї мають найменше ознак перетренованості (overtraining).

3.4 Тестування роботи додатку

Для тестування роботи додатку буде виконано дві генерації з однаковими параметрами та текстовим описом, перша генерація не адаптованою мережею, друга – адаптованою. В результаті друга генерація повинна бути максимально наближена до початкових референсів на основі яких було виконано адаптацію.



Рисунок 3.4 – Порівняння результатів

Рис.3.4 має наступний текстовий опис: a frozen cosmic spritz soda can, the petals glitter with a crystalline shimmer, swirling nebulas, 8k unreal engine photorealism, ethereal lighting, red, nighttime, darkness, surreal art.

Негативний запит: pixels, watermark, bad hands, low quality, noise.

Steps: 30

CFG scale: 7

Seed: 3109513708

Size: 1024x1024

Рис.3.5 згенерований за текстовим описом: ethereal fantasy concept art of a spritz soda can, magnificent, celestial, ethereal, painterly, epic, majestic, magical, fantasy art, cover art, dreamy.

Негативний запит: pixels, watermark, bad hands, low quality, noise.

Steps: 35

CFG scale: 7

Seed: 4040829509

Size: 1024x1024



Рисунок 3.5 – Порівняння результатів

Судячи з отриманих результатів, зроблено висновок що адаптація пройшла успішно та додаток працює вірно. Отримані адаптовані результати максимально схожі на початкові референси.

ВИСНОВКИ

У ході виконання кваліфікаційної роботи бакалавра було виконано низку важливих етапів для успішної реалізації проекту з розроблення програмного додатку нейромережевої генерації зображень для рекламних кампаній.

Проведено аналіз сучасних досліджень та технологій у сфері нейромережевої генерації зображень. Розглянуто такі потужні моделі, як Stable Diffusion, DALL-E та Midjourney, а також перспективні методи адаптації моделей, зокрема Low-Rank Adaptation (LoRA), Textual Inversion та DreamBooth. Обґрунтовано актуальність створення програмного додатку для генерації зображень для рекламних кампаній з використанням цих технологій.

Здійснено аналіз програмних аналогів, порівнявши DALL-E 3, Midjourney та Stable Diffusion з точки зору їх можливостей, особливостей та придатності для вирішення поставленої задачі. Зроблено висновок, що Stable Diffusion є найбільш оптимальним вибором завдяки відкритому вихідному коду та можливості персоналізації моделі за допомогою описаних методів.

Досягнуто мету проекту – створення програмного додатку для генерації зображень за допомогою нейромережевих технологій, призначеного для використання у рекламних кампаніях. Додаток має надавати зручний інтерфейс для роботи з моделями генерації зображень.

Визначено та детально описано завдання, які необхідно вирішити для досягнення поставленої мети, включаючи розробку технічного завдання, вибір оптимальної моделі, створення функціоналу та графічного інтерфейсу, тестування та підготовку документації.

Складено технічне завдання (Додаток А), що містить детальний опис вимог до програмного додатку, його структури, етапів розробки та інших аспектів.

Виконано планування робіт (Додаток Б), визначено необхідні ресурси та проаналізовано можливі ризики реалізації проекту.

Успішно реалізовано проект з розробки програмного додатку для генерації зображень для рекламних кампаній з використанням нейромережевих технологій.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. High-Resolution Image Synthesis with Latent Diffusion Models. 13 квіт. 2022 р. / Robin Rombach та ін. arXiv. URL: <https://arxiv.org/abs/2112.10752> (дата звернення: 30.04.2024).
2. Midjourney Documentation and User Guide. URL: <https://docs.midjourney.com/> (дата звернення: 30.04.2024).
3. LoRA: Low-Rank Adaptation of Large Language Models. 16 жовт. 2021 р. / E. J. Hu та ін. arXiv. URL: <https://arxiv.org/abs/2106.09685> (дата звернення: 23.04.2024).
4. An Image is Worth One Word: Personalizing Text-to-Image Generation using Textual Inversion. 2 серп. 2022 р. / R. Gal та ін. arXiv. URL: <https://arxiv.org/abs/2208.01618> (дата звернення: 23.04.2024).
5. Utilizing stable diffusion and fine-tuning models in advertising production and logo creation: An application of text-to-image technology. 2024 р. / Chenyang Wang. URL: <https://ace.ewapublishing.org/media/1d8df9cfe72543809429345ebc4e0341.marked.pdf> (дата звернення 30.04.2024)
6. DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation. 15 берез. 2023 р. / Nataniel Ruiz та ін. arXiv. URL: <https://arxiv.org/abs/2208.12242> (дата звернення: 23.04.2024).
7. AI-Generated Content (AIGC): A Survey. 26 берез. 2023 р. / Wensheng Gan та ін. arXiv. URL: <https://arxiv.org/abs/2304.06632> (дата звернення: 30.04.2024)
8. Improving Image Generation with Better Captions. 2023 р. / James Betker та ін. URL: <https://cdn.openai.com/papers/dall-e-3.pdf> (дата звернення: 30.04.2024)
9. Libraries - OpenAI API. URL: <https://platform.openai.com/docs/libraries/community-libraries> (дата звернення: 30.04.2024)
10. Image generation - OpenAI API. URL: <https://platform.openai.com/docs/guides/images/usage> (дата звернення: 01.05.2024)

11. stabilityai/stable-diffusion-xl-base-1.0 Hugging Face. URL: <https://huggingface.co/stabilityai/stable-diffusion-xl-base-1.0> (дата звернення: 01.05.2024)
12. PyTorch. URL: <https://pytorch.org/> (дата звернення: 01.05.2024)
13. DALL·E: Creating images from text. 5 січн. 2021 р. / Aditya Ramesh та ін. URL: <https://openai.com/research/dall-e> (дата звернення: 30.04.2024)
14. Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding. 23 трав. 2022 р. / Jonathan Ho та ін. arXiv. URL: <https://arxiv.org/abs/2205.11487> (дата звернення: 30.04.2024)
15. Image generation - OpenAI API. URL: <https://platform.openai.com/docs/guides/images> (дата звернення 30.04.2024)
16. Malik, M. E. та ін. Impact of brand image and advertisement on consumer buying behavior. World Applied Sciences Journal. 2013 р. Vol. 23. P. 117-122.
17. A survey of Generative AI Applications. 14 черв. 2023 р. / Eduardo C. Garrido-Merchán та ін. arXiv. URL: <https://arxiv.org/abs/2306.02781> (дата звернення: 30.04.2024)
18. Dena Magdy Hanna. The Use of Artificial Intelligence Art Generator "Midjourney" in Artistic and Advertising Creativity. Journal of Design Sciences and Applied Arts. черв. 2023 р. Vol. 4. P. 42-58.
19. SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis. 4 липн. 2023 р. / Robin Rombach та ін. arXiv. URL: <https://arxiv.org/abs/2307.01952> (дата звернення: 30.04.2024)
20. Wo-Kyeong Kim, Hyoung-Kyu Song, Thibault Castells, та Shinkook Choi. ВК-SDM: A Lightweight, Fast, and Cheap Version of Stable Diffusion. arXiv:2305.15798, 2023.
21. NovelAI. Novelai improvements on stable diffusion, 2023. URL: <https://blog.novelai.net/novelai-improvements-on-stable-diffusion-e10d38db82ac> (дата звернення: 30.04.2024)
22. Aditya Ramesh. How dall·e 2 works, 2022. URL: <http://adityaramesh.com/posts/dalle2/dalle2.html> (дата звернення: 30.04.2024)

23. Lvmin Zhang та Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. arXiv:2302.05543, 2023.

24. Kohya's GUI. URL: https://github.com/bmaltais/kohya_ss (дата звернення: 21.05.2024)

25. Jonathan Ho та Tim Salimans. Classifier-Free Diffusion Guidance. arXiv:2207.12598, 2022.

ДОДАТОК А.

ТЕХНІЧНЕ ЗАВДАННЯ

**на розробку програмного додатку
«Програмний додаток нейромережевої генерації
зображень для рекламних кампаній»**

ПОГОДЖЕНО:

Доцент кафедри інформаційних технологій

_____ Парфененко Ю.В.

Студент групи ІТ-03-1

_____ Волков П.К.

Суми 2024

1. Призначення й мета створення програмного додатку

1.1 Призначення програмного додатку

Програмний додаток призначений для автоматизованої генерації зображень за допомогою нейромережових моделей для використання у рекламних кампаніях.

1.2 Мета створення програмного додатку

Метою створення програмного додатку є надання користувачам зручного інструменту для швидкої та ефективної генерації зображень, які можна використовувати у рекламних кампаніях. Додаток базується на навчених моделях глибокого навчання, таких як Stable Diffusion, здатних створювати якісні та стилізовані зображення.

1.3 Цільова аудиторія

Цільовою аудиторією програмного додатку є маркетологи, рекламні агентства, дизайнери та інші фахівці, які займаються рекламними кампаніями та потребують швидкої генерації якісних зображень.

2 Вимоги до програмного додатку

2.1 Вимоги до програмного додатку в цілому

2.1.1 Вимоги до структури й функціонування програмного додатку

Додаток повинен мати інтуїтивно зрозумілий інтерфейс для користувачів.

Функціонал додатку повинен включати можливість введення текстового запиту для генерації зображень. Приклад текстового запиту: “moody aesthetic, beautiful cozy, cramped bedroom with floor to ceiling glass windows overlooking a cyberpunk city at night, view from top of skyscraper, white bedsheets, bookshelves, thunderstorm outside with torrential rain, detailed, high resolution, photorealistic, dark, gloomy”.

Додаток повинен підтримувати підключення різних контрольних точок нейромережевої моделі Stable Diffusion для генерації зображень, наприклад: “Juggernaut XL”, “Anything XL” та інші сумісні XL моделі.

Наявність додаткових налаштувань для контролю параметрів генерації зображень, таких як: Seed, CFG Scale, Sampling steps.

Генерування зображень з роздільною здатністю від 512*512 до 1024*1024 пікселів.

2.1.2 Вимоги до персоналу

Відсутність вимог до спеціалізованого персоналу. Додаток повинен бути доступним для використання користувачами з різним рівнем кваліфікації.

2.1.3 Вимоги до збереження інформації

Додаток повинен забезпечувати безпечне збереження всієї інформації, включаючи введений текстовий запит, параметри генерації та згенеровані зображення.

Всі дані користувачів, включаючи текстові запити та згенеровані зображення, повинні бути збережені локально на пристрої користувача з можливістю експорту або імпорту.

2.2 Структура програмного додатку

2.2.1 Загальна інформація про структуру програмного додатку

Програмний додаток складатиметься з графічного інтерфейсу, який буде взаємодіяти з основними функціями генерації зображень за допомогою нейромережевих алгоритмів.

2.2.2 Навігація

Навігація в додатку повинна бути легкою та зрозумілою для користувачів. Наявність основних вкладок або розділів, таких як введення текстового запиту, вибір моделі та налаштування.

2.2.3 Дизайн та структура додатку

Дизайн інтерфейсу повинен бути зручним для користувача та відповідати сучасним стандартам UI/UX дизайну. Потрібно передбачити окремі вкладки або панелі для введення текстових запитів, вибору параметрів генерації, та перегляду результатів.

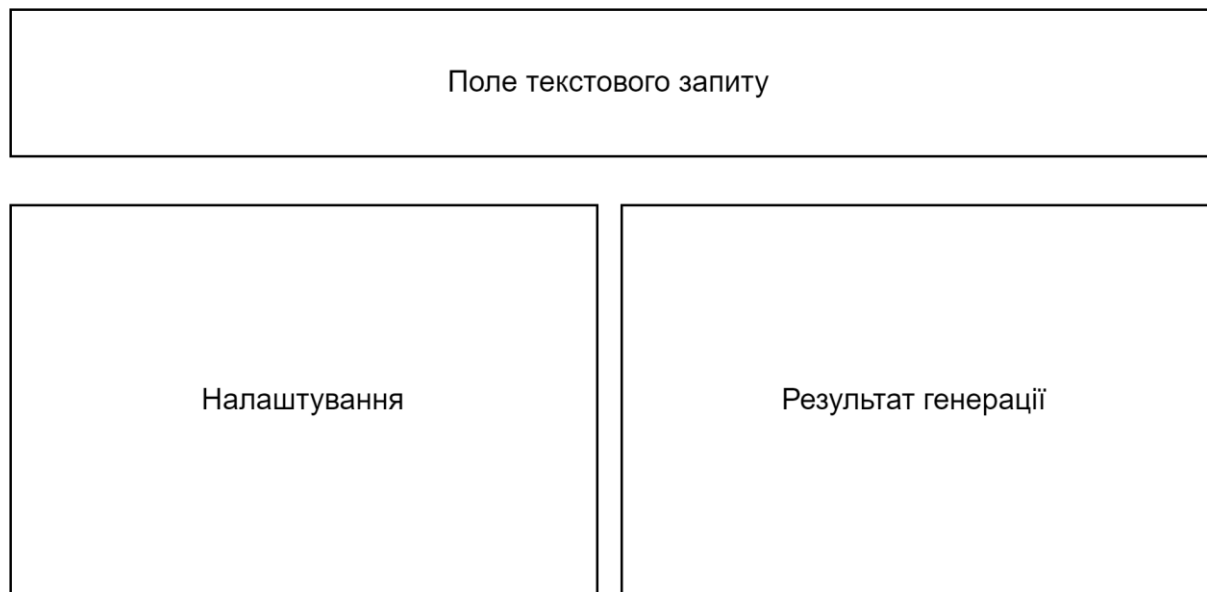


Рисунок А.1 – Схематичне зображення інтерфейсу

2.3 Вимоги до функціонування системи

2.3.1 Потреби користувача

Потреби користувача представлені у таблиці А.1.

Таблиця А.1 – Потреби користувача

ID	Потреби користувача	Джерело
UN-01	Можливість введення текстового запиту для генерації зображень	Користувач
UN-02	Налаштування параметрів генерації зображень	Користувач
UN-03	Перегляд результатів генерації зображень	Користувач
UN-04	Збереження згенерованих зображень	Користувач

2.3.2 Функціональні вимоги

На основі потреб користувача були визначені такі функціональні вимоги:

- можливість вводити текстові запити для генерації зображень.
- налаштування параметрів генерації зображень.
- перегляд результатів генерації зображень в реальному часі.
- можливість зберігання згенерованих зображень на пристрої користувача.

2.3.3 Системні вимоги

Даний розділ визначає, розподіляє та вказує на системні вимоги, визначені розробником. Їх перелік наведений в таблиці А.2.

Таблиця А.2 – Системні вимоги

ID	Системні вимоги	Пріоритет	Опис
SR-01	Підтримка операційних систем: Windows, macOS, Linux	М	Додаток повинен бути сумісний з операційними системами Windows, macOS та Linux.
SR-02	Python версії 3.10	М	Для коректної роботи додатку потрібна встановлена версія Python 3.10.
SR-03	Оперативна пам'ять 8Гб	М	Мінімально рекомендована кількість оперативної пам'яті для оптимальної роботи додатку - не менше 8 Гб
SR-04	Наявність GPU з 8ГБ VRAM	М	Додаток вимагає наявності GPU з мінімальною кількістю 8 Гб VRAM для генерації зображень.
SR-05	15Гб вільного місця на накопичувачі	М	Мінімально рекомендована кількість вільної пам'яті на накопичувачі - не менше 15 Гб.

Умовні позначення в таблиці А.2:

Must have (M) – вимоги, які повинні бути реалізовані в системі;

2.4 Вимоги до видів забезпечення

2.4.1 Вимоги до інформаційного забезпечення

Додаток повинен надавати документацію, яка включає в себе інструкції з використання та опис основних функцій.

2.4.2 Вимоги до лінгвістичного забезпечення

Програмний додаток має бути виконаний англійською мовою.

3 Склад і зміст робіт зі створення програмного додатку

Докладний опис етапів роботи зі створення програмного додатку наведено в таблиці А.3.

Таблиця А.3 – Етапи створення програмного додатку

№	Склад і зміст робіт	Строк розробки (у робочих днях)
1	Постановка цілей необхідних для досягнення певного результату	1 день
2	Складання технічного завдання	3 дні
3	Підготовка референсних зображень	5 днів
4	Навчання нейромережі	3 дні
5	Тестування працездатності нейромережі	3 дні
6	Пошук оптимальних параметрів	2 дні
7	Розробка графічного інтерфейсу	7 днів
8	Перевірка працездатності програмного додатку	3 дні
9	Завершення роботи	2 дні
	Загальна тривалість робіт	29 днів

ДОДАТОК Б. ПЛАНУВАННЯ РОБІТ

Деталізація мети проекту методом SMART. Продуктом кваліфікаційної роботи бакалавра є програмний додаток нейромережевої генерації зображень для рекламних кампаній.

Результати деталізації методом SMART розміщені у табл. Б.1.

Таблиця Б.1 – Деталізація мети методом SMART

Specific (конкретна)	Розробка програмного додатку, який буде здатен генерувати зображення на основі навченої нейромережі
Measurable (вимірювана)	Перевірка успішності завершення проекту здійснюється через оцінку функціональності та якості згенерованих зображень, а також задоволеність користувачів результатами.
Achievable (досяжна)	Мета досяжна через використання наявних технологій у галузі машинного навчання та розробки програмного забезпечення.
Relevant (реалістична)	Урахування обмежень технічних можливостей і ресурсів для реалізації програмного продукту.
Time-framed (обмежена у часі)	Програмний додаток повинен бути розроблений протягом календарного плану.

Планування змісту структури робіт. Основним інструментом для планування змісту структури робіт служить WBS діаграма (Work Breakdown Structure), яка допомагає розбити проект на окремі фази та завдання для ефективного управління проектом. Діаграма WBS зображена на рис. Б.1.

Планування структури організації, для впровадження готового проекту (OBS). Після побудови WBS розробимо організаційну структуру виконавців OBS

(Organizational Breakdown Structure). Діаграма OBS зображена на рис. Б.2. Список виконавців, що функціонують в проекті, знаходиться в табл. Б.2.

Таблиця Б.2 – Виконавці проекту

Роль	Ім'я	Проектна роль
Розробник	Волков П.К.	Займається розробкою основного функціоналу проекту та створенням інтерфейсу користувача
Тестувальник	Волков П.К.	Відповідає за тестування функціоналу, а також проводить перевірку моделі на відповідність вимогам.

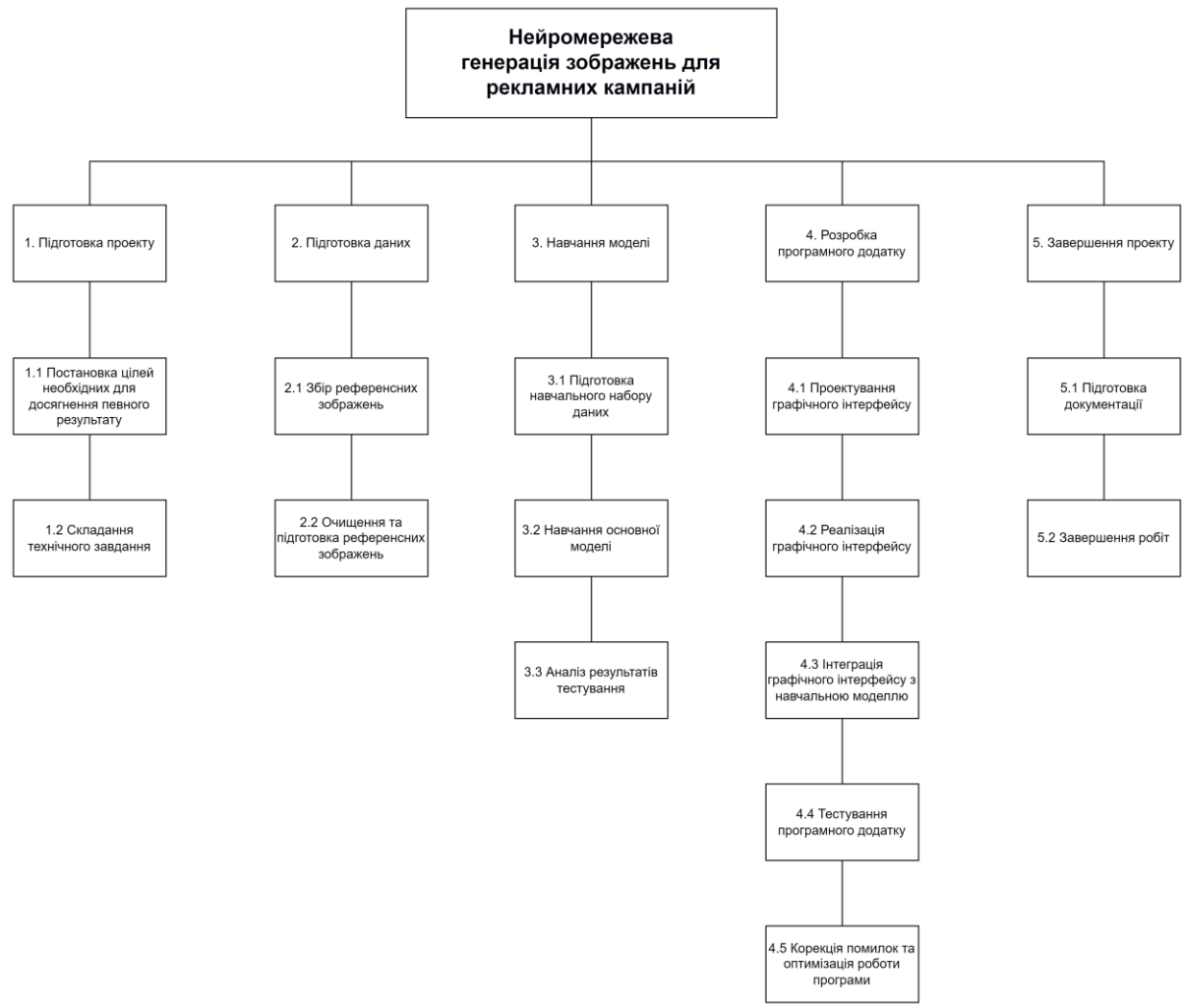


Рисунок Б.1 – WBS. Структура робіт проекту

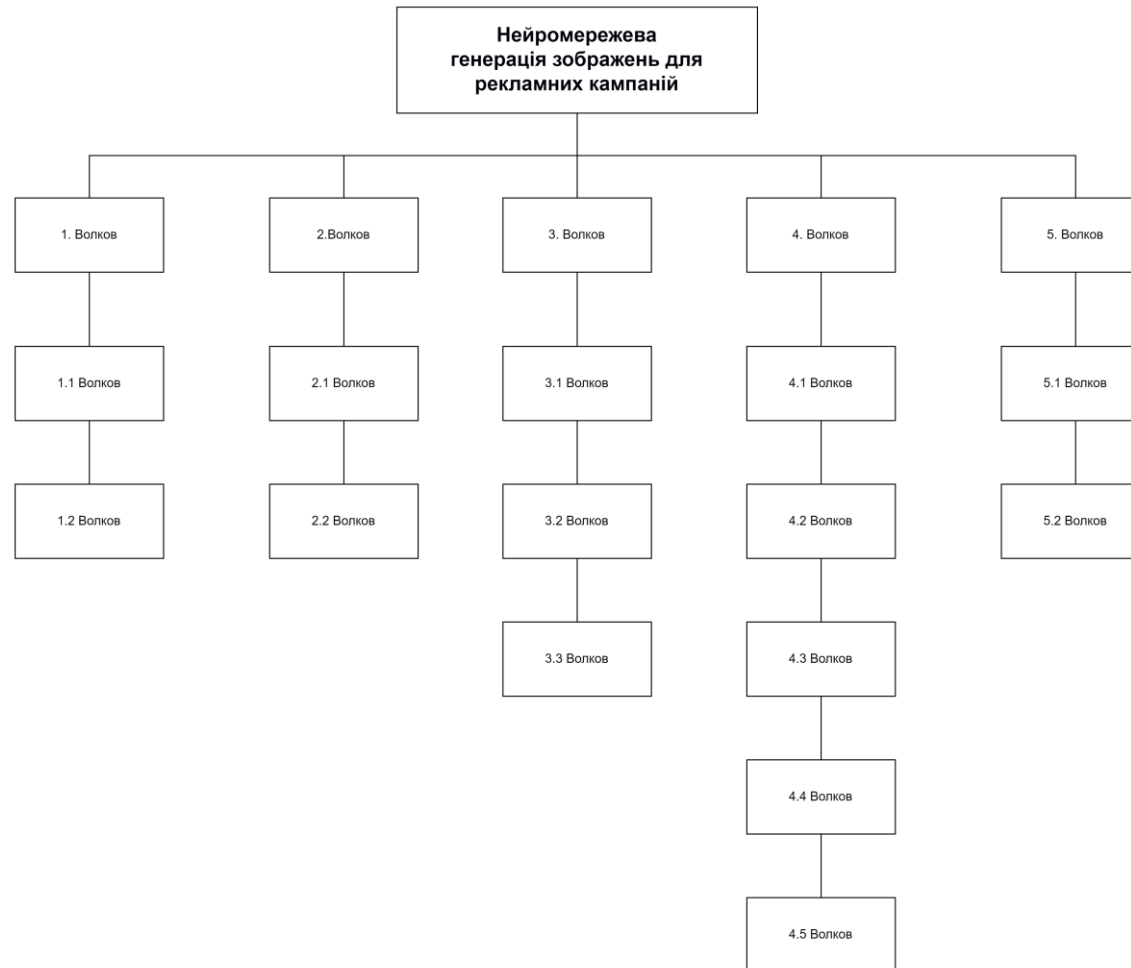


Рисунок Б.2 – Організаційна структура проекту (OBS)

Діаграма Ганта. Для планування виконання дипломного проекту буде використана Діаграма Ганта. Цей інструмент є основним у більшості проектних управлінських практик. Діаграма Ганта надає можливість візуалізувати часові рамки та послідовність завдань, що дозволяє команді розробників краще управляти проектом. Тривалість робіт буде вказана в днях. Для більш точного уявлення про тривалість робіт, з урахуванням обмежень та ресурсів, буде побудовано календарний графік. Діаграма Ганта та список робіт у ній зображені на рисунку Б.3.

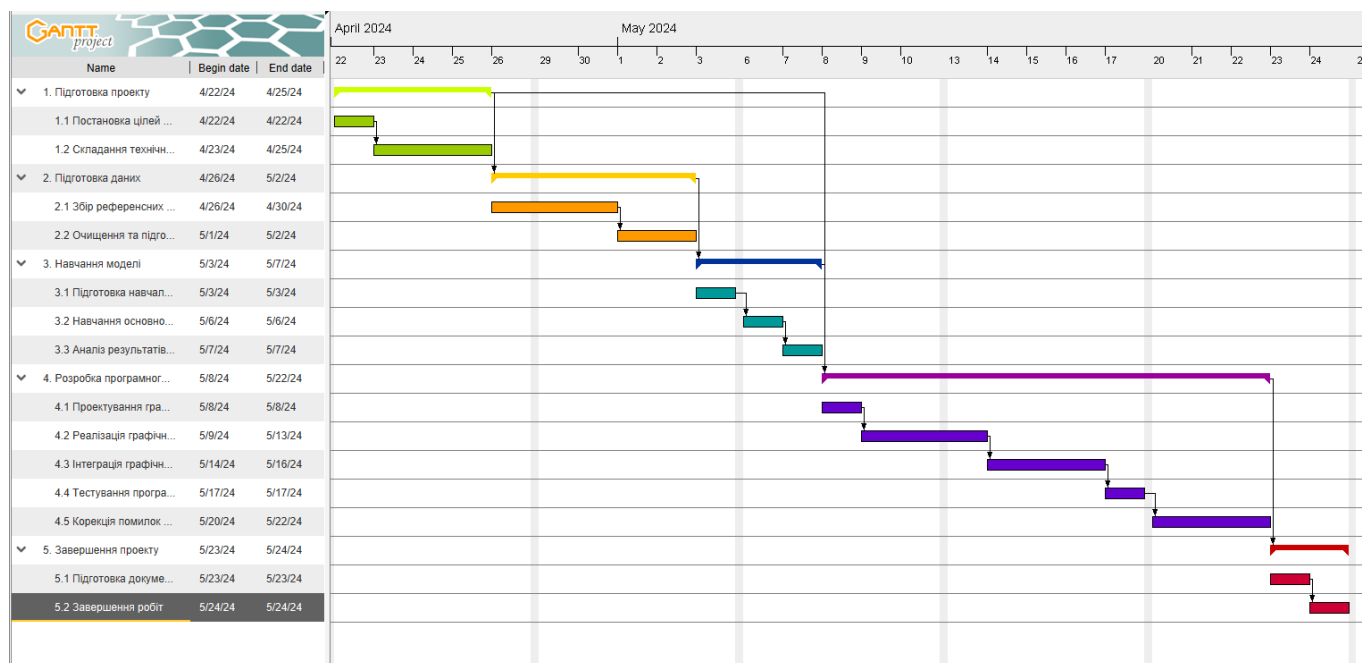


Рисунок Б.3 – Діаграма Ганта

Аналіз ризиків. У рамках проекту буде проведений аналіз якісних та кількісних аспектів ризиків. Під час якісної оцінки будуть ідентифіковані найбільш значущі ризики, що потребують негайного реагування. Кількісний аналіз ризиків дозволить отримати більш повне уявлення про їх вплив на проект. Обидва підходи можуть використовуватися окремо або разом, залежно від обсягу ресурсів та необхідності. У таблиці Б.5 наведена класифікація ризиків за ймовірністю виникнення та величиною втрат.

Після аналізу ризиків буде розроблено план реагування на них, що включає в себе розробку методів та технологій мінімізації негативного впливу. Також буде визначено, чи можуть наслідки впливу ризику на проект бути позитивними чи негативними. Оцінка ризиків за параметрами з таблиці Б.3 дозволить побудувати матрицю ймовірності виникнення та впливу ризиків, яка буде представлена на рисунку Б.4.

Таблиця Б.3 – Шкала оцінювання ймовірності виникнення та впливу ризику на виконання проекту

Оцінка	Ймовірність виникнення	Вплив ризику
1	Низька	Низький
2	Середня	Середній
3	Висока	Високий

Ймовірність виникнення	3			
	2			RS_2
	1	RS_3	RS_1	RS_4, RS_5
		1	2	3
	Вплив ризику			

Рисунок Б.4 – Матриця ймовірності виникнення ризиків та впливу ризику

- зелений колір – прийнятні ризики;
- жовтий колір – виправданні ризики;
- червоний колір – недопустимі ризики.

На підставі отриманого значення індексу ризику класифікують: за рівнем ризику, що знаходиться в табл. Б.4.

Таблиця Б.4 – Шкала оцінювання за рівнем ризику

№	Назва	Межі	Ризики, які входять(номера)
1	Прийнятні	$1 \leq R \leq 2$	1,3
2	Виправдані	$3 \leq R \leq 4$	2,5
3	Недопустимі	$6 \leq R \leq 9$	4

Таблиця Б.5 – Оцінка ймовірності виникнення, величини витрат та індексу ризику

ID	Статус ризику	Опис ризику	Ймовірність виникнення	Вплив ризику	Ранг ризику	План А	Тип стратегії реагування	План Б
RS_1	Відкритий	Невдалий процес навчання нейромережі	Низька	Середній	1	Перегляд параметрів навчання, вдосконалення референсних зображень	Попередження	Мінімізація використання LoRA, вивчення альтернативних методів
RS_2	Відкритий	Низька якість вихідних зображень після навчання	Середня	Високий	3	Тюнінг параметрів генерації, використання більшої кількості референсів	Попередження	Пошук інших генеративних методів, які можуть покращити якість
RS_3	Відкритий	Проблеми з оптимізацією та швидкістю процесу генерації	Низька	Низький	2	Оптимізація коду	Прийняття	Розгляд можливості використання більш потужних обчислювальних ресурсів
RS_4	Відкритий	Непередбачувані труднощі у взаємодії з іншими інструментами та середовищами	Низька	Високий	6	Співпраця з розробниками інших інструментів	Прийняття	Розробка альтернативних інтерфейсів для інших інструментів
RS_5	Відкритий	Низька стабільність результатів генерації при використанні різних текстових запитів	Низька	Високий	3	Систематичне тестування з різними текстовими запитами	Попередження	Вивчення методів покращення стабільності при різних запитах

ДОДАТОК В. ЛІСТНИГ ПРОГРАМНОГО КОДУ

```
import random
from PIL import Image
from io import BytesIO
import base64
import requests
import gradio as gr

SD_FOLDER = "./stable-diffusion-webui"

def generate_images(prompt_negative, prompt, cfg_scale, steps, width, height,
seed_value, model_selection, lora_selection):
    prompt = f"{prompt} <lora:{lora_selection}:1>"
    sd_parameters = {
        "prompt": prompt,
        "negative_prompt": prompt_negative,
        "seed": seed_value,
        "sampler_name": "Euler a",
        "steps": steps,
        "cfg_scale": cfg_scale,
        "width": width,
        "height": height,
    }

    response = requests.post("http://127.0.0.1:7860/sdapi/v1/txt2img",
json=sd_parameters)

    if response.status_code == 200:
        response = response.json()
        image = response["images"][0]
        image = Image.open(BytesIO(base64.b64decode(image)))

        if seed_reuse.value == False:
            seed_value = randomize_seed()

        return image, seed_value
    else:
        print(response.json()["detail"]["msg"])
        return None, seed_value

def randomize_seed():
    return random.randint(0, 9000000000)

sd_models = requests.get("http://127.0.0.1:7860/sdapi/v1/sd-models").json()
sd_loras = requests.get("http://127.0.0.1:7860/sdapi/v1/loras").json()
```

```

with gr.Blocks(
    title="Stable Diffusion Image Generation",
    theme=gr.themes.Soft(),
) as demo:
    with gr.Row():
        with gr.Column(scale=3):
            with gr.Row():
                model_selection = gr.Dropdown([model["model_name"] for model in
sd_models],
                    label="Model Selection",
                    value=sd_models[0]["model_name"],
                    filterable=False,
                    allow_custom_value=False,
                    scale=1,
                )
                lora_selection = gr.Dropdown([lora["name"] for lora in sd_loras],
                    label="LoRa Selection",
                    filterable=False,
                    allow_custom_value=False,
                    scale=2,
                )

            with gr.Column(scale=4):
                pass
        with gr.Row():
            with gr.Column():
                prompt_positive = gr.Textbox(label="Positive Prompt", lines=3)
                prompt_negative = gr.Textbox(label="Negative Prompt", lines=3)
            with gr.Column():
                image_output = gr.Image(show_download_button=True)
                generate_btn = gr.Button(value="Generate", variant="primary")
    gr.Markdown("## Generation Settings")
    with gr.Row(variant="compact"):
        with gr.Column():
            steps = gr.Slider(1, 100, label="Steps", value=20, step=1, info="Number
of steps to take in the diffusion process")
            cfg_scale = gr.Slider(0.1, 10, label="CFG Scale", value=7, step=0.05)
        with gr.Row():
            seed = gr.Number(label="Seed", value=randomize_seed(), step=1,
scale=4)

            seed_rnd = gr.Button(value="Randomize", scale=1, size="sm")
            seed_reuse = gr.Checkbox(label="Reuse Seed", scale=1)
        with gr.Column():
            width = gr.Number(label="Width", value=1024, step=1)
            height = gr.Number(label="Height", value=1024, step=1)

    generate_btn.click(generate_images,
        [prompt_negative,
        prompt_positive,
        cfg_scale,
        steps,
        width,

```

```
        height,  
        seed,  
        model_selection,  
        lora_selection,  
        ], outputs=[image_output, seed])  
seed_rnd.click(randomize_seed, outputs=seed)  
  
demo.launch(server_port=7861)
```

ДОДАТОК Г. АПРОБАЦІЯ РЕЗУЛЬТАТІВ ДОСЛІДЖЕННЯ

ІМА :: 2024

СЕКЦІЯ 2: Інформаційні технології проектування

Програмний додаток генерації візуального контенту за допомогою нейронних мереж

Волков П. К., студент ІТ-03-1; Парфененко Ю. В., доцент

Сумський державний університет, м. Суми, Україна

Сучасні генеративні моделі штучного інтелекту, такі як Stable Diffusion, відкривають нові горизонти для створення унікального візуального контенту. Stable Diffusion XL 1.0 – це потужна нейромережа, здатна генерувати високоякісні зображення з роздільною здатністю 1024x1024 пікселів. Метод LoRA (Low-Rank Adaptation) дозволяє адаптувати модель для генерації зображень з унікальними об'єктами чи в певному стилі, забезпечуючи безпрецедентну гнучкість та креативність.

Метою даної роботи є розробка програмного додатку, який інтегрує навчену за допомогою LoRA нейромережу Stable Diffusion XL 1.0 для генерації візуального контенту для рекламних кампаній. Цей додаток стане потужним інструментом для маркетингових відділів компаній, які прагнуть створювати привабливі та унікальні візуальні матеріали для просування своїх продуктів чи послуг.

Для ефективного використання цієї технології необхідно подолати певні технічні бар'єри. По-перше, потрібно підготувати набір високоякісних зображень-референсів для навчання нейромережі. По-друге, процес навчання вимагає значних обчислювальних ресурсів, зокрема відеопам'яті обсягом понад 20 ГБ, що часто потребує оренди потужних серверів з відеоприскорювачами. Програмний додаток нейромережевої генерації зображень інтегруватиме Stable Diffusion XL 1.0 з метою генерування візуального контенту відповідно до вимог рекламних кампаній. Він забезпечить зручний інтерфейс для налаштування параметрів моделі, генерації зображень та управління отриманими результатами.

Розробка програмного додатку відкриє нові можливості для маркетингових команд в створенні привабливого та унікального візуального контенту, який допоможе ефективно просувати їхні продукти чи послуги. Застосування передових технологій генеративного штучного інтелекту дозволить компаніям підвищити креативність та якість своїх рекламних кампаній, залишаючись на передовій в умовах жорсткої конкуренції.