

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Національний університет «Запорізька політехніка»

Факультет комп'ютерних наук та технологій

(повне найменування факультету)

Кафедра комп'ютерних систем та мереж

(повне найменування кафедри)

Пояснювальна записка

до дипломного проекту (роботи)

бакалавра

(ступінь вищої освіти)

на тему Розробка методики генерації зображень з використанням
нейронних мереж
(назва теми)

Виконав(ла): студент(ка) 4 курсу, групи КНТ 510 сп

Спеціальності 123 «Комп'ютерна інженерія»

(код і найменування спеціальності)

Освітня програма (спеціалізація)

«Комп'ютерна інженерія»

СЕМЕЛЬЯНОВ Д. В.

(ПРИЗВИЩЕ та ініціали)

Керівник ГОЛУБ Т.В.

(ПРИЗВИЩЕ та ініціали)

Рецензент КУЛЯБА-ХАРИТОНОВА Т.І

(ПРИЗВИЩЕ та ініціали)

2023

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Національний університет «Запорізька політехніка»
(повне найменування вищого навчального закладу)

Факультет _____ Комп'ютерних наук і технологій
Кафедра _____ «Комп'ютерні системи та мережі»
Ступінь вищої освіти _____ бакалаврський
Спеціальність _____ 123 Комп'ютерна інженерія
(код і найменування)
Освітня програма (спеціалізація) _____ Комп'ютерна інженерія
(назва освітньої програми (спеціалізації))

ЗАТВЕРДЖУЮ

Завідувач кафедри Кудерметов Р.К.

“ _____ ” _____ 2023 року

З А В Д А Н Н Я
НА ДИПЛОМНИЙ ПРОЄКТ (РОБОТУ) СТУДЕНТА

СЕМЕЛЬЯНОВА Дмитра Владиславовича

(ПРІЗВИЩЕ, ім'я, по батькові)

1. Тема проєкту (роботи) Розробка методики генерації зображень з використанням нейронних мереж

керівник проєкту (роботи) к. т. н., доцент, ГОЛУБ Тетяна Василівна

(науковий ступінь, вчене звання, ПРІЗВИЩЕ, ім'я, по батькові)

затверджені наказом вищого навчального закладу від “24” березня 2023 року № 71

2. Строк подання студентом проєкту (роботи) _____ 23 травня 2023 року

3. Вихідні дані до проєкту (роботи) _____ наявні сучасні системи генерації зображень з використанням нейронних мереж, методичні вказівки що до їх використання, порівняння та аналіз отриманих результатів, а також надання рекомендацій стосовно використання

4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити)

1) Огляд нейронних мереж та аналіз інструментів генерації зображень;

2) Генерація зображень з використанням нейронних мереж;

3) Порівняння та аналіз роботи НМ генерації зображень.

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень, кількість слайдів плакатів)

Порівняння згенерованих зображень різних НМ (А3);

Додаткові інструменти НМ Midjourney (А3);

Додаткові інструменти НМ Stable Diffusion (А3);

Додаткові інструменти НМ DALL·E 2 (А3).

6. Консультанти розділів проєкту (роботи)

Розділ	ПРІЗВИЩЕ, ініціали та посада консультанта	Підпис, дата	
		завдання видав	прийняв виконане завдання
1-3	ГОЛУБ Т.В., к.т.н., доцент		
нормоконтроль	ЩЕРБАК Н.В., ст. викл.		

7. Дата видачі завдання 01.03.2023 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів дипломного проєкту (роботи)	Строк виконання етапів проєкту (роботи)	Примітка
1	Робота з джерелами інформації	01.03.2023 р.	
2	Вибір нейронних мереж які будуть розглядатися	08.03.2023 р.	
3	Робота з документацією	12.03.2023 р.	
4	Генерація зображень використовуючи Midjourney	31.03.2023 р.	
5	Генерація зображень використовуючи Stable Diffusion	13.04.2023 р.	
6	Генерація зображень використовуючи DALL-E 2	29.04.2023 р.	
7	Оформлення отриманих результатів у ПЗ	07.05.2023 р.	
8	Оформлення графічного матеріалу	18.05.2023 р.	
9	Оформлення допоміжного матеріалу	22.05.2023 р.	

Студент (ка)


(підпис)

Дмитро СЕМЕЛЬЯНОВ
(Ім'я ПРІЗВИЩЕ)

Керівник проєкту (роботи)


(підпис)

Тетяна ГОЛУБ
(Ім'я ПРІЗВИЩЕ)

РЕФЕРАТ

ПЗ: 89 с., 100 рис., 6 табл., 22 джерел.

CLIP, PROMPT, МОДЕЛЬ, НЕЙРОННА МЕРЕЖА, ПІДСКАЗКА,
ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ, СИСТЕМА, ШТУЧНИЙ ІНТЕЛЕКТ

Об'єкт розробки – методики генерації зображень з використанням нейронних мереж.

Предмет дослідження – генерація зображень з використанням нейронних мереж.

Метою роботи є розробка методики генерації зображень з використанням нейронних мереж.

Проект складається з трьох розділів.

Перший розділ присвячено огляду та опису предметної області сучасних систем, які дозволяють генерувати зображення, використовуючи нейронні мережі. Наведено опис принципів дії та методів штучного інтелекту, та вказівки для початку роботи з нейронними мережами.

У другому розділі розкрито основні положення генерації зображень з використанням нейронних мереж, обрано теми для генерації зображень та розглянуто безпосередній процес генерації.

У третьому розділі проведено порівняння та аналіз отриманих у другому розділі результатів.

В результаті виконання дипломної роботи було сформовано методику генерації зображень різних стилістичних напрямків з використанням нейронних мереж.

ЗМІСТ

Скорочення та умовні позначки	8
Вступ.....	9
1 Огляд нейронних мереж та аналіз інструментів для генерації зображень.....	10
1.1 Сучасні системи генерації зображень	12
1.1.1 Midjourney	13
1.1.2 DALL-E.....	13
1.1.3 Stable Diffusion.....	14
1.1.4 Adobe Firefly	15
1.1.5 Craiyon	16
1.2 Програмні засоби генерації зображень	17
1.2.1 Midjourney	18
1.2.2 Stable Diffusion.....	21
1.2.3 DALL-E 2.....	26
2 Генерація зображень з використанням нейронних мереж.....	30
2.1 Основні теоретичні положення генерації зображень	30
2.2 Складання початкової мети для аналізу та порівняння	31
2.3 Робота із Midjourney.....	31
2.3.1 Відмінності у версіях на прикладі кота	32
2.3.2 Будівля.....	33
2.3.3 Людина	34
2.3.4 Пейзаж	38
2.3.5 Текстура.....	40

2.3.6 Автомобіль	43
2.3.7 Листівка	44
2.3.8 Розгляд інших інструментів	45
2.4 Робота із Stable Diffusion	49
2.4.1 Відмінності у версіях на прикладі кота	50
2.4.2 Будівля	52
2.4.3 Людина	55
2.4.4 Пейзаж	59
2.4.5 Текстура	62
2.4.6 Автомобіль	64
2.4.7 Листівка	65
2.4.8 Розгляд інших інструментів	66
2.5 Робота із DALL-E 2	68
2.5.1 Відмінності у версіях на прикладі кота	68
2.5.2 Будівля	68
2.5.3 Людина	69
2.5.4 Пейзаж	71
2.5.5 Текстура	72
2.5.6 Автомобіль	73
2.5.7 Листівка	75
2.5.8 Розгляд інших інструментів	77
3 Порівняння та аналіз роботи НМ для генерації зображень	79
3.1 Відмінності у версіях	79
3.2 Робота із Midjourney	81
3.3 Робота із Stable Diffusion	82

3.4 Робота із DALL-E 2	83
3.5 Рекомендації та методика використання НМ.....	84
Висновки	86
Перелік джерел посилання	88

Графічний матеріал:

Плакат 1 - Порівняння згенерованих зображень різних НМ

Плакат 2 - Додаткові інструменти НМ Midjourney

Плакат 3- Додаткові інструменти НМ Stable Diffusion

Плакат 4 - Додаткові інструменти НМ DALL·E 2

СКОРОЧЕННЯ ТА УМОВНІ ПОЗНАЧКИ

НМ – Нейронна мережа

ШІ – Штучний інтелект

ПЗ – Програмне забезпечення

ML – Machine Learning

ВСТУП

На сьогоднішній день розвиток нейронних мереж (НМ) досяг достатньо високого рівня. Розроблено нейронні мережі для вирішення задач багатьох різноманітних напрямів. При цьому використовуються різні технології їх реалізації.

НМ непогано справляється з аналізом даних, їх сортуванням, оптимізацією, прогнозуванням. Майже з кожним днем НМ знаходить використання в різних галузях від вирішення певних вузькоспрямованих задач промислового масштабу до пошуку відповідей на побутові питання ентузіастів.

Одними із напрямків використання НМ, який викликає все більшу зацікавленість у виконанні різноманітних задач є розпізнавання та обробка зображень. Напрямок має доволі широке застосування.

Так, використання автоматичного енкодера, побудованого на основі згорткових НМ, дозволяє видалити шуми (таким інструментом, наприклад, є DLSS від NVIDIA). Також камери багатьох сучасних смартфонів використовують НМ для виконання фільтрації зображень.

Окреме місце в роботі із зображеннями займає процес їх генерації. Дипломний проект присвячений дослідженню існуючих нейронних мереж, що спеціалізуються на генерації зображень. А також розробці рекомендацій щодо обрання критеріїв і інструментів для генерації зображень різних стилістичних напрямків.

1 ОГЛЯД НЕЙРОННИХ МЕРЕЖ ТА АНАЛІЗ ІНСТРУМЕНТІВ ДЛЯ ГЕНЕРАЦІЇ ЗОБРАЖЕНЬ

Коли заходить мова про НМ у застосуванні до зображень, маються на увазі НМ що розпізнають, задача яких полягає в аналізі зображення, спочатку дивляться на значення пікселів зображення, застосовують до картинки набір перетворень і видають інформацію, наприклад, про клас об'єкта (що саме на ній зображено).

В останні роки все більше уваги присвячено НМ, які вирішують обернені завдання, тобто завдання синтезу зображень. Зворотне завдання полягає в наступному: комп'ютеру необхідно переглянути велику кількість зображень, що містять певні об'єкти, суб'єкти, та, вивчивши, що таке наприклад обличчя або будинок, а закінчивши навчання НМ повинна вчитися створювати нові зображення.

Однією з важливих задач при цьому є те, що утворені зображення мають бути, по-перше, коректними, з погляду людини; по-друге, мають істотно відрізнятися від тих об'єктів, суб'єктів, які нейромережа «бачила» в процесі навчання.

Наприклад для генерації нових зображень обличч нейронна мережа повинна визначати, що таке обличчя, краще, ніж для розпізнавання. Розпізнавати обличчя НМ вже зараз вміє в багатьох умовах не гірше, ніж людина. Система, яка синтезує зображення обличч, просто «подивившись» перед цим набір інших зображень обличч, не досягає особливо хороших результатів. Особи виходять або розмиті, або на них є суттєві артефакти та спотворення, або ці особи виходять надто схожими на ті, які комп'ютер бачив у процесі навчання. З математичної точки зору в даній задачі потрібно якимось чином явно чи неявно моделювати ймовірнісний розподіл у багатовимірному просторі, в якому «існують» зображення. Причому моделювати їх так, щоб точки відповідні реальному зображенню, мали б більшу ймовірність, ніж інші.

Майже усі підходи вирішують це завдання багатовимірним розподілом (вибір не такий великий - гауссівський розподіл у багатовимірному просторі, або

рівномірний розподіл на сфери). Далі НМ навчається, викривляє багатовимірний простір, ускладнюючи його, наближаючи його до опису очікуваного зображення. Припустимо, нам вдалося знайти таку НМ, яка правильним чином викривляє простір, і тоді ми можемо синтезувати зображення дуже просто: ми беремо наш початковий рівномірний розподіл, витягаємо з нього крапку, пропускаємо її через НМ і на виході отримуємо синтезоване зображення з обличчям, будівлею або твариною.

Концептуально все досить просто, але проблема в реалізації. Це область математичної статистики. Останніми роками дослідники у сфері нейромереж повернулися до розробки чи, можливо, до перевідкриття класичних методів математичної статистики. Робота з такими високорозмірними уявленнями вимагає залучення методів і теорії ймовірності, і диференціальної геометрії, і функціонального аналізу. Значна частина сучасної математики у цій галузі дуже затребувана.

Один з підходів, в який багато дослідників дуже вірять, це підхід, пов'язаний з протиборчими мережами (adversarial networks). Він полягає в тому, що до навчання генеруючої мережі, яка викривляє простий розподіл у потрібний нам розподіл, навчається ще одна мережа – дискримінуюча (оцінююча). Процес навчання полягає у грі, тобто спостерігається залучення теорії «ігор».

В результаті генерації зображень за допомогою НМ можуть бути створені варіанти об'єктів які здатні стати основою для їх подальшого фізичного втілення, вже почали з'являтися приклади втілення речей, згенерованих штучним інтелектом, в реальному житті. Прикладом є відтворення меблів, створених НМ. Також доцільним може бути використання згенерованих за допомогою НМ зображень для оформлення обкладинки музичних альбомів, розробки оригінальних текстур та в інших сферах дизайну.

Даний дипломний проект полягає у дослідженні роботи різних існуючих НМ для генерації зображень, порівнянні отриманих результатів та зрештою розробці методики генерації зображень певних стилістичних типів, а також проведенні аналізу можливостей НМ та особливостей роботи із ними.

1.1 Сучасні системи генерації зображень

З появою нових досліджень та технологій продовжується розвиток НМ. Вже на даний момент існує багато різних їх варіантів, частина з яких приведена у таблиці 1.1

Таблиця 1.1 - Сучасні НМ генерації зображень

НМ генерації зображень	Короткі відомості	Цінова політика
Midjourney	Найпоширеніша НМ генерації зображень	Платний доступ
Bing Image Creator	Виконана на базі DALL-E, але з безкоштовною	Безкоштовний доступ
DALL·E 2	Дуже інтуїтивна та швидка НМ з якою можна досягти гарних результатів	Платний доступ
StarryAI	генератор мистецтва, який дозволяє перетворювати мистецтво в NFT	Платний доступ
Adobe Firefly	НМ для людей які працюють з редагуванням контенту	Безкоштовний доступ
Deep Dream Generator	Генерує надмірно оброблені зображення посиляючись на Парейдолою.	Безкоштовний доступ
НМ генерації зображень	Короткі відомості	Цінова політика
Stable Diffusion	НМ з відкритим вихідним кодом, з великою гнучкістю	Безкоштовний доступ
Craiyon	Похідна нейромережа від DALL·E	Умовно безкоштовний
MyHeritage's AI Time Machine	Генератор власних портретів	Безкоштовний

До найпопулярніших систем які досягли непоганих результатів в цьому напрямку, відносяться: Midjourney, Stable Diffusion, DALL-E, Adobe Firefly, Craiyon.

1.1.1 Midjourney

Система, яка складається зі служби та програми генеративного штучного інтелекту, створена та розміщена незалежною дослідницькою лабораторією Midjourney, Inc. у Сан-Франциско. Midjourney генерує зображення з описів природною мовою, які називаються «prompts», схожа за принципами роботи до OpenAI DALL-E та Stable Diffusion [1], [2].

Дана НМ найбільш художня серед інших. Інструмент наразі перебуває у відкритому бета-тестуванні, яке було запущено 12 липня 2022 року. Лідером команди Midjourney є Девід Хольц, який є також співзасновником Leap Motion.

Користувачі створюють ілюстрації використовуючи Midjourney за допомогою команд бота у Discord.

Дана НМ найбільш художня серед інших. Інструмент наразі перебуває у відкритому бета-тестуванні, яке було запущено 12 липня 2022 року. Лідером команди Midjourney є Девід Хольц, який є також співзасновником Leap Motion.

Користувачі створюють ілюстрації, використовуючи Midjourney, за допомогою команд бота у Discord.

Компанія працює над вдосконаленням своїх алгоритмів, випускаючи нові версії моделей кожні кілька місяців. Версія №2 їхнього алгоритму була запущена в квітні 2022 року, а версія №3 — 25 липня. 5 листопада 2022 року для користувачів було випущено альфа-ітерацію версії №4, а 15 березня 2023 року було випущено альфа-ітерацію версії №5. Модель №5.1 є більш «переконливою», ніж версія №5, застосовуючи більше власної стилізації до зображень, тоді як модель 5.1 RAW додає покращення, але працює краще з більш буквальними «підказками».

1.1.2 DALL-E

DALL-E і DALL-E 2 — це моделі глибокого навчання, розроблені OpenAI для створення цифрових зображень із описів природною мовою, які називаються «prompts». DALL-E було опубліковано OpenAI у дописі в блозі в січні 2021 року, і

він використовує версію GPT-3 [3], модифіковану для створення зображень. У квітні 2022 року OpenAI анонсувала DALL-E 2, наступника, призначеного для створення більш реалістичних зображень із вищою роздільною здатністю, які «можуть поєднувати концепції, атрибути та стилі» [4].

OpenAI не опублікував вихідний код для обох моделей. 20 липня 2022 року DALL-E 2 увійшов у бета-фазу із запрошеннями, надісланими 1 мільйону осіб зі списку очікування; користувачі можуть створювати певну кількість зображень безкоштовно щомісяця та можуть купувати більше.

Раніше доступ був обмежений для попередньо відібраних користувачів для попереднього перегляду дослідження через занепокоєння щодо етики та безпеки. 28 вересня 2022 року DALL-E 2 було відкрито для всіх, а вимогу щодо списку очікування було скасовано.

Цікавий факт, назва програмного забезпечення є набір імен мультиплікаційного персонажа Pixar WALL-E та іспанського художника-сюрреаліста Сальвадора.

1.1.3 Stable Diffusion

Це модель глибокого навчання для перетворення тексту в зображення, випущена в 2022 році. Вона в основному використовується для створення детальних зображень на основі текстових описів, хоча її також можна застосовувати для виконання інших завдань, таких як замальовування та створення зображень. Розроблена стартапом Stability AI у співпраці з рядом академічних дослідників і некомерційних організацій [5].

Stable Diffusion — це модель не вимірної дифузії, різновид глибокої генеративної нейронної мережі. Її код і ваги моделі знаходяться у вільному доступі, та вона може працювати на більшості споживчого обладнання, оснащеного скромним графічним процесором із принаймні 8 ГБ відеопам'яті. Це ознаменувало відхід від попередніх власних моделей перетворення тексту в зображення, таких як DALL-E та Midjourney, які були доступні лише через хмарні служби.

Публічний реліз Stable Diffusion змусив бити на сполох деяких людей, які побоюються його культурних та економічних наслідків. На відміну від DALL-E 2,

тренувальні дані Stable Diffusion ("ваги") доступні будь-кому для використання без будь-яких суворих обмежень.

Офіційний реліз Stable Diffusion (i Dream Studio) включають автоматичний фільтр NSFW (наготи) і додають невидимий термінг для трекінгу, але ці обмеження можна легко обійти, коли код відкритий для всіх. Це означає, що Stable Diffusion можна використовувати для створення зображень, які OpenAI зараз блокує в DALL-E 2: пропаганда, жорстокі зображення, порнографія, зображення з потенційним порушенням авторських прав, дипфейки знаменитостей тощо. Заради справедливості, ліцензія Stable Diffusion офіційно забороняє багато з цих використань, але коли код і ваги знаходяться у загальному доступі, проконтролювати це вкрай важко, якщо взагалі можливо. У відповідь на ці побоювання Мостак сказав, що на його думку, переваги відкритого доступу до такого інструменту (коли його можна вивчати) переважають недоліки [6].

1.1.4 Adobe Firefly

Adobe Firefly — це сімейство моделей штучного інтелекту для набору програм і служб Adobe, зокрема штучного інтелекту для створення медіаконтенту. Firefly — це розширення попередніх генеративних інструментів штучного інтелекту, представлених Adobe для Photoshop, Express і Lightroom під час конференції Max 2022. Firefly буде інтегровано безпосередньо в робочі процеси Adobe Creative Cloud, Document Cloud, Experience Cloud і Adobe Express, ставши частиною серії нових генеративних служб Adobe Sensei AI. Adobe стверджує, що моделі штучного інтелекту навчаються на вмісті, ліцензованому або захищеному авторським правом, а не на роботах художників, знайдених в Інтернеті.

Adobe представила Firefly 21 березня 2023 року, випустивши бета-версію, орієнтовану на комерційних користувачів, із двома початковими інструментами. Перший дозволяє користувачам вводити підказки природною мовою та отримувати зображення, подібно до інших генеративних моделей зображень AI, таких як DALL-E та Midjourney. Другий генерує стилізований текст, застосовуючи стилі або текстури до літер і шрифтів. На момент запуску бета-версія Firefly доступна лише

на веб-сайті, але Adobe заявила, що незабаром вона буде інтегрована в її програми, включаючи Express, Photoshop, Illustrator і Adobe Experience Manager [7].

Перші моделі Firefly навчалися на сотнях мільйонів зображень Adobe Stock, вмісті з відкритою ліцензією та вмісті, що є суспільним надбанням. Firefly прагне не створювати на основі чужих брендів чи інтелектуальної власності. Adobe планує запровадити тег «Не тренуватися», щоб дозволити творцям відмовитися від використання їх вмісту для навчання моделей. Adobe також прагне дати клієнтам можливість монетизувати свої творіння, розробивши модель винагороди для творців Adobe Stock [8].

Виходячи з можливостей та особливостей Adobe Firefly стає дуже потужним та зручним інструментом для людей працюючих у сфері створення контенту й дизайну, але поки вільного доступу на момент створення дипломного проекту немає, він розширюється по системі подачі заявок та поступовому прийомі користувачів.

1.1.5 Craiyon

Craiyon — це генератор зображень зі штучним інтелектом (AI), здатний генерувати зображення з текстових «підказок». Craiyon було розроблено як полегшену версію DALL-E від OpenAI і розроблено, щоб бути таким же простим у використанні, як і оригінальний DALL-E. Раніше відомий як DALL-E Mini, Craiyon пізніше змінив назву, щоб уникнути плутанини з моделями DALL-E та DALL-E 2 від OpenAI, а також на прохання OpenAI.

Компанію заснував у 2022 році Борис Дайма, штаб-квартира знаходиться в Х'юстоні, штат Техас. Заснування відбулося після того, як Борис Дайма взяв участь у JAX/Flax Community Week у червні 2021 року та вперше запропонував концепцію Craiyon на форумах Hugging Face. Початкова версія навчалася протягом трьох днів на одному TPUv3-8, наданому Google Cloud протягом усього терміну дії програми. Незважаючи на обмеження в часі та обчислювальній архітектурі, це була в основному успішна концепція.

Craiyon - це безкоштовний інструмент для некомерційних цілей - із платними рівнями підписки для тих, хто цікавиться комерційним використанням - і

спонсорується за рахунок реклами для безкоштовного рівня. Платні рівні також пропонують менший час очікування для створення зображень.

Спочатку модель Craiyon навчалася на мільйонах зображень з Інтернету та супроводжуючих їх підписах. Ці підписи спонукають модель вибирати зображення для використання на основі текстових підказок. Модель навчили поєднувати концепції для генерації нових зображень із підказки, на яку вона має посилання [9].

1.2 Програмні засоби генерації зображень

Кожна описана НМ має свої особливості використання, що зумовлює деякі відмінності при виборі та користуванні ними.

Craiyon хоч і перетворився на окремий продукт, все ж таки є похідним від DALL-E, а до Adobe Firefly не усі одразу можуть отримати доступ та він є специфічним серед розглянутих НМ тому вони будуть представлені в детальному розгляді. Особливу увагу приділено першим трьом із перелічених вище.

Робота НМ із зображеннями представляє собою трудомісткий процес, розглянуті Midjourney та DALL-E 2 виконуються на серверах правовласників і це зручно в певному плані, але для роботи із ними необхідно буде сплачувати кошти за використання.

Stable Diffusion, в свою чергу, має відкритий вихідний код та поширюється на безкоштовній основі, тому для роботи з нею можна використовувати власну робочу станцію, або навіть персональний комп'ютер. Також існує багато онлайн сервісів, де можна працювати з НМ за різними варіантами, аналогічно іншим на платних умовах чи безкоштовних з обмеженнями, які полягають в гнучкості та функціональності мережі. При роботі на власному обладнанні доступні усі налаштування, а найголовніше – це можливість використовувати різні моделі в самій НМ.

1.2.1 Midjourney

Ця НМ дуже проста у знайомстві, для отримання певного, відносно точного результату не потрібно значних знань, але для більш конкретного результату необхідно ознайомитися з документацією та особливостями значень, що вводяться, але це стосується не тільки даної НМ.

Із документацією можна ознайомитися на офіційному сайті у розділі посібника (рис. 1.1), який можна знайти за посиланням [10].

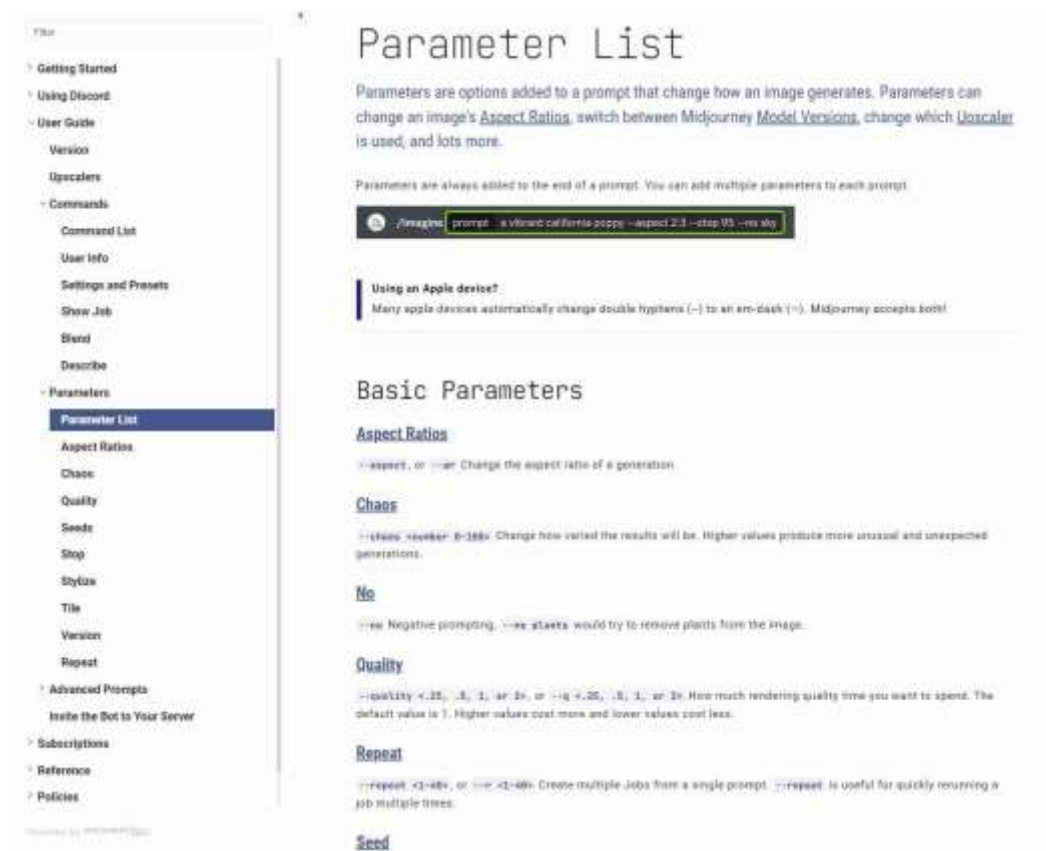


Рисунок 1.1 - Посібник Midjourney

Midjourney - це інструмент для створення зображень зі штучним інтелектом, який отримує вхідні дані через текстові підказки та параметри та використовує алгоритм машинного навчання ML, навчений на великій кількості даних зображень, щоб створювати унікальні зображення.

Наразі Midjourney доступний лише через бота Discord на офіційному Discord. Користувач створює зображення за допомогою команди «/imagine» і вводить у

командний рядок, як і будь-який інший інструмент для генерації зображень використовуючи нейронні мережі. Потім бот повертає знімок.

Взаємодія із системою відбувається через discord. Та для цього необхідно мати обліковий запис, і якщо його немає - зареєструватися. Взаємодія між користувачем і системою відбувається за допомогою спеціально створеного для цього бота. Інтерфейс для взаємодії наведено на рисунку 1.2.

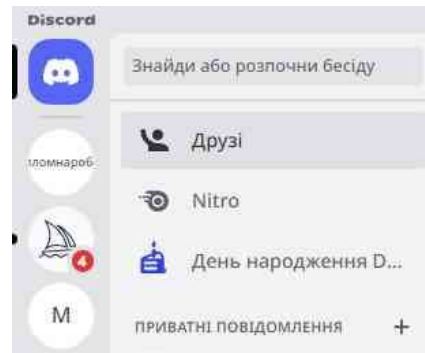


Рисунок 1.2 - інтерфейс discord

Щоб розпочати роботу з НМ необхідно виконати ряд послідовних кроків.

Першим кроком необхідно перейти на офіційний сайт Midjourney за посиланням [11] та на самому сайті натиснути «приєднатися» що створить посилання на discord канал. Там знаходяться публічні чати де можна подивитись корисну інформацію або вже безпосередньо зображення, які генерують різні користувачі в реальному часі. Також на цьому кроці вже можна попередньо оцінити НМ основуючись на побачених у каналі прикладах.

Другим кроком є початок роботи. Така можливість надається за умови наявності ліцензії у користувача. Генерація виконується з використанням бота в офіційному discord каналі midjourney. Для цього необхідно перейти у будь-який зі спеціально виділених для цього чатів (рис.1.3), або ж його можна додати на власний канал для зручності використання чи інших потреб. Реалізацію процесу створення власного каналу та залучення до нього боту приведено на рисунках 1.3 – 1.5.

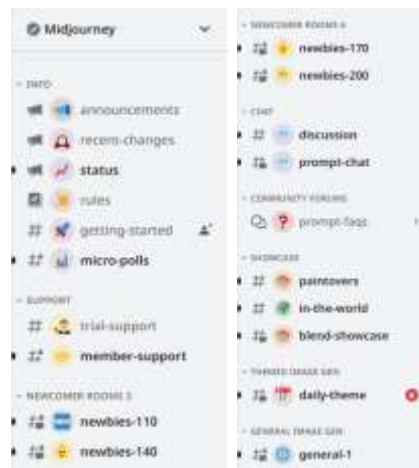


Рисунок 1.3 - Перелік чатів офіційного каналу midjourney

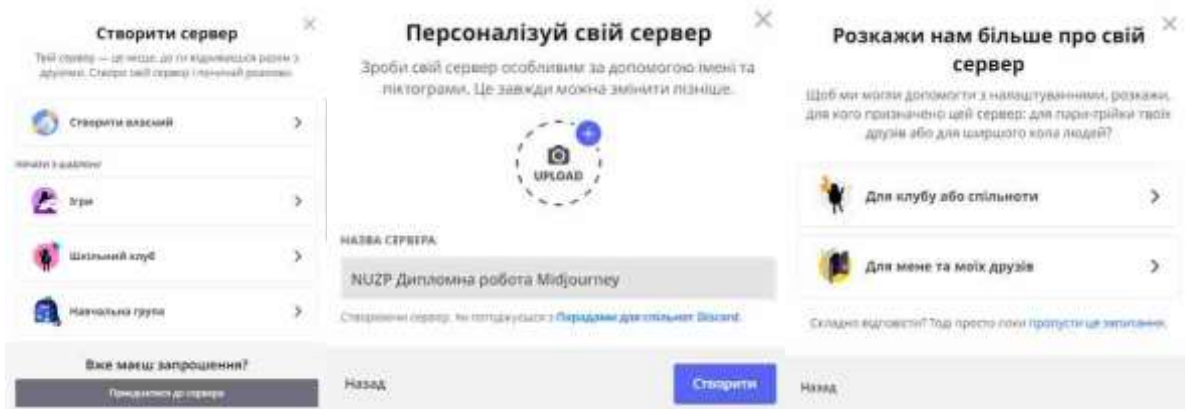


Рисунок 1.4 - Створення власного каналу у discord

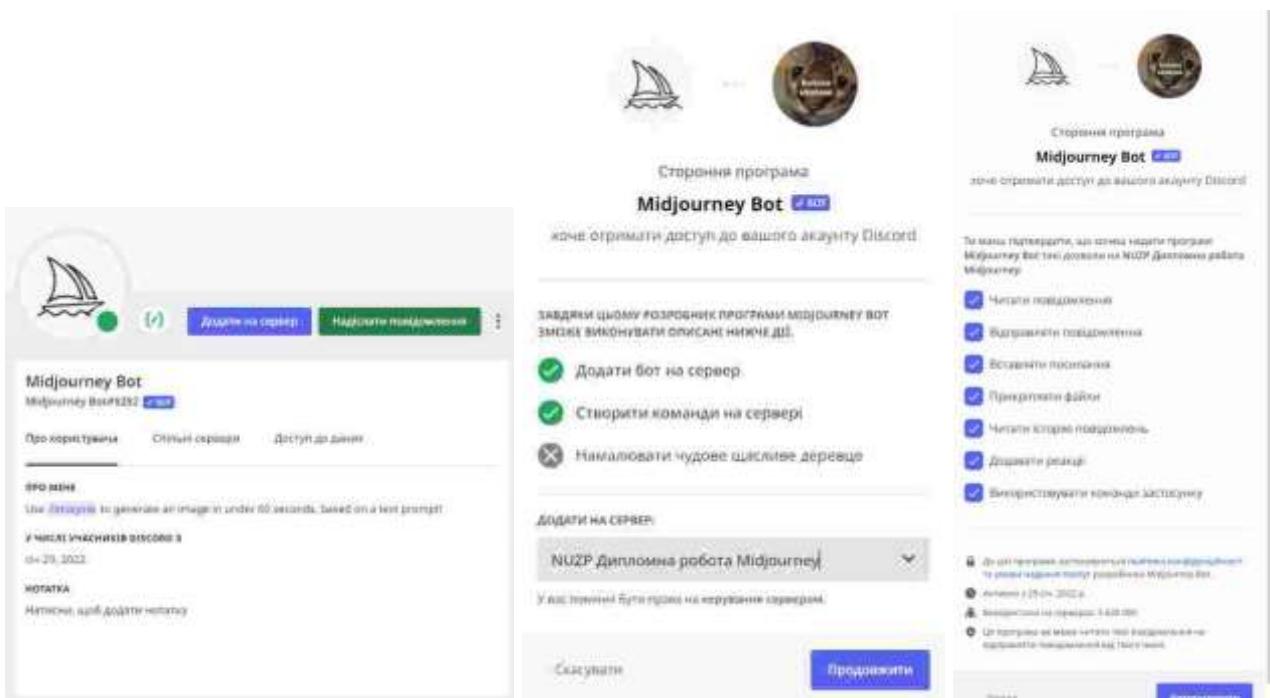


Рисунок 1.5 - Додавання боту midjourney на власний канал

Midjourney має закритий вихідний код та поширюється на платній основі. Усі операції виконуються на заздалегідь вже готових серверах та з боку користувача необхідні тільки вхідні дані, та за це доводиться платити.

Користування даною НМ не є безкоштовним, варіанти планів приведені на рис.1.6. Базовий план підходить для знайомства, але для розширеного та менш обмеженого використання доведеться придбати стандартний план.

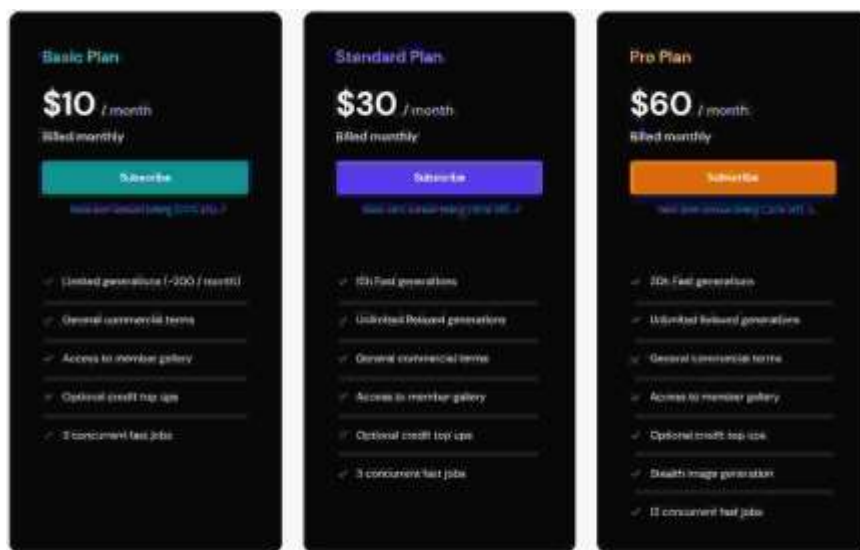


Рисунок 1.6 - Перелік планів покупки Midjourney

1.2.2 Stable Diffusion

Має відкритий вихідний код, та будь-хто з бажаючих може використовувати його і, за потреби, навіть встановити на власну робочу станцію. Відкритість також дає доступ до усіх внутрішніх налаштувань, що робить дану НМ більш гнучкою у цьому пункті, але потребує більшого досвіду використання та підвищує поріг входу.

Особливостями даної НМ є наступні.

Для створення картинок Stable Diffusion використовує набір даних LAION-5B, що містить 5 мільярдів зображень з інтернету, включаючи популярні сайти, такі як Pinterest, DeviantArt та Getty Images, але для користувацьких моделей з будь якого джерела набір даних буде відрізнятися.

Stable Diffusion — це модель перетворення тексту в зображення, яка використовує «заморожений» текстовий кодувальник CLIP (Contrastive Language–Image Pre-training) ViT-L/14 для налаштування моделі за текстовими підказками.

НМ CLIP розроблена OpenAI. CLIP був розроблений для пом'якшення низки основних проблем у стандартному підході глибокого навчання до комп'ютерного «зору».

Метод використовує доступне джерело контролю: текст у поєднанні з зображеннями, знайденими в Інтернеті. Ці дані використовуються для створення наступного навчального завдання проксі-сервера для CLIP: надання зображення, передбачення, який із набору з 32 768 випадково відібраних фрагментів тексту фактично відноситься з ним у нашому наборі даних.

Вирішення завдання полягає в тому, що моделям CLIP необхідно буде навчитися розпізнавати широкий спектр візуальних концепцій у зображеннях і пов'язувати їх із їхніми назвами. У результаті моделі CLIP можна застосовувати до майже довільних завдань візуальної класифікації. Наприклад, якщо завданням набору даних є класифікація фотографій собак і котів, ми перевіряємо для кожного зображення, чи прогнозує модель CLIP текстовий опис «фото собаки» або «фото кота» з більшою вірогідністю бути парним. з цим, для наглядності приведено рисунках 1.7 – 1.8.

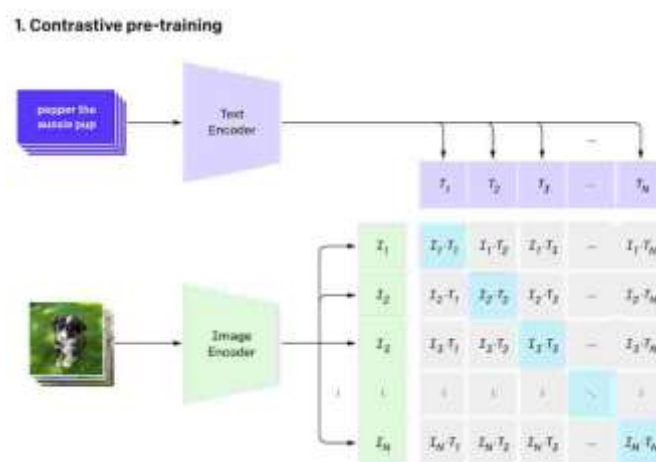


Рисунок 1.7 – Зіставне попереднє тренування

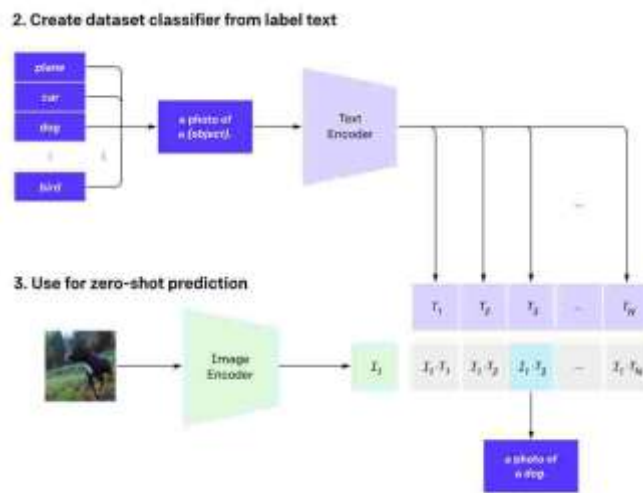


Рисунок 1.8 – створення датасету класифікаторів з тексту

CLIP попередньо навчає кодувальник зображень і кодувальник тексту, задля передбачення, які зображення були поєднані з яким текстом у нашому наборі даних. З рештою ця поведінка використовується, щоб перетворити CLIP на нульовий класифікатор. Усі класи набору даних перетворюються у підписи, такі як «фото собаки», та прогнозується клас підпису. CLIP оцінює найкращі пари з певним зображенням.

Детальніша інформація про технологію CLIP наведено за посиланням [11].

Stable Diffusion використовує CLIP для розподілу процесу створення зображення на процес «розповсюдження» під час виконання — він починається лише з шуму та поступово покращує зображення, доки воно повністю не звільниться від шуму, поступово наближаючись до наданого текстового опису.

Завдяки «відкритості» має різні моделі які можуть створюватися користувачами ентузіастами, що безпосередньо впливає на результат.

Stable Diffusion не має офіційної документації стосовно «prompt», але існує так званий Stable Diffusion prompts search engine [12], який представляє собою пошукову систему, що у відповідь на введений «prompt» видає результати з використанням даного тегу. Також задля розуміння принципу використання «prompt» існує неофіційний «Prompt Book», в якому знаходиться більшість корисної інформації стосовно використання тексту [13].

Щоб розпочати роботу з НМ необхідно виконати ряд послідовних кроків.

Першим кроком потрібно встановити необхідне ПЗ, актуальні версії Python та Git.

Наступним кроком необхідно встановити веб інтерфейс. Його можна знайти на GitHub за посиланням [14].

Передостаннім кроком треба завантажити необхідну модель із сайту huggingface за посиланням [15]. Цей сайт присвячений зберіганню різноманітних моделей та інструментів, пов'язаних з ШІ, в даному випадку це stable-diffusion-v1-5. Після завантаження файл з описом моделі помістити у теку моделей Stable Diffusion (рис. 1.9).

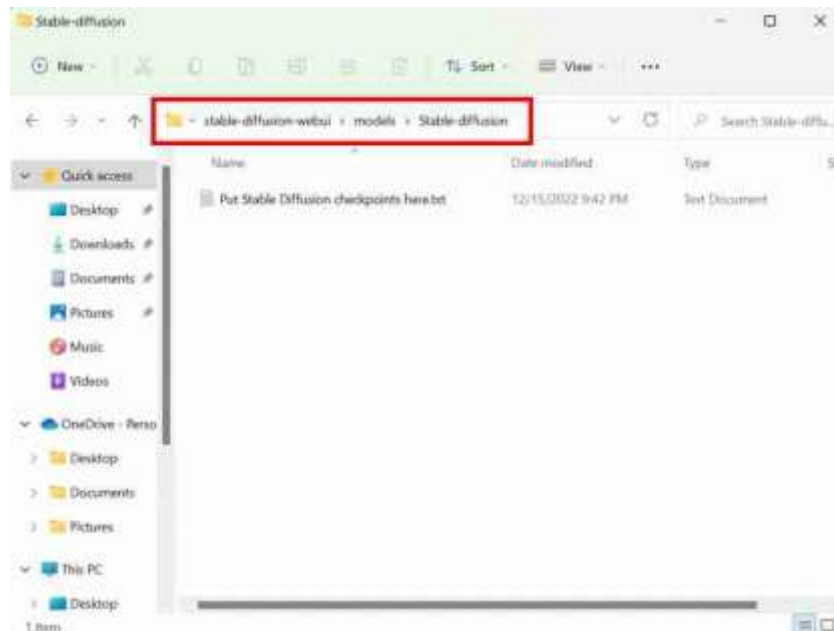


Рисунок 1.9 - Шлях розташування моделей в теці

Останнім кроком є запуск самого застосунку. Для цього необхідно запустити .bat файл, зразок приведено на рис. 1.10. Відкриється командний рядок, де можна побачити процес запуску. При першому встановленні застосунок автоматично дозавантажить необхідні інструменти. Також у даному командному рядку будуть відображатися всі виконувані процеси (рис. 1.11).

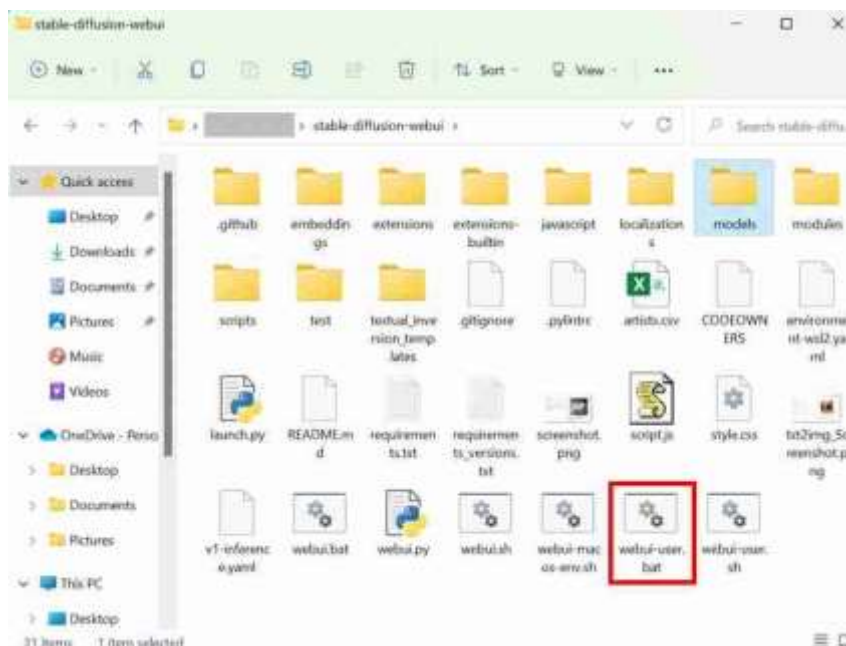


Рисунок 1.10 – Шлях розташування виконавчого файлу в теці

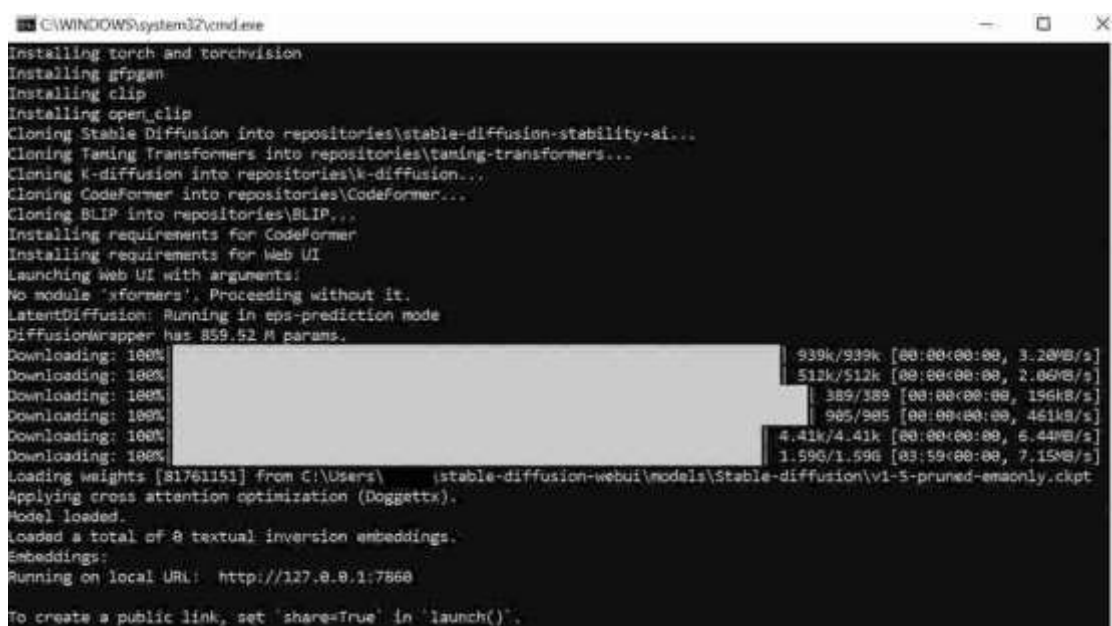


Рисунок 1.11 – Командний рядок запуску застосунку

Після виконання усіх попередніх кроків можна переходити до безпосередньої роботи з НМ за допомогою веб інтерфейсу на який можна потрапити за адресою, яка виводиться у командному рядку. Стандартний інтерфейс застосунку приведено на рис. 1.12.

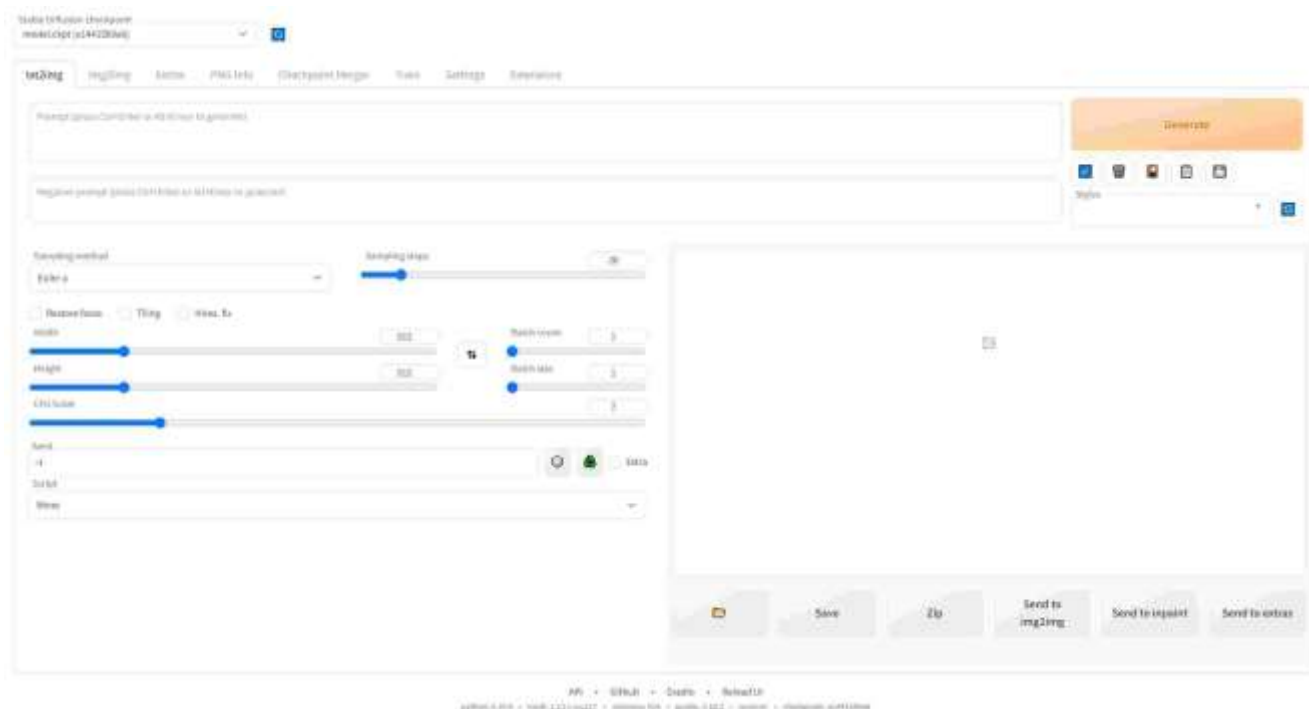


Рисунок 1.12 - веб інтерфейс застосунку Stable Diffusion

1.2.3 DALL-E 2

DALL-E 2 аналогічно до Midjourney має закритий вихідний код, та створює зображення з текстових описів. Використовуючи понад 10 мільярдів параметрів тренувальних версій моделі трансформатора GPT-3, він інтерпретує вхідні дані природною мовою «prompt» та генерує відповідне зображення.

Щоб розпочати роботу з НМ необхідно виконати ряд послідовних кроків.

DALL-E 2 в першу чергу складається з 2 частин: одна для перетворення введених користувачем даних у представлення зображення (Prior), а інша - для перетворення цього подання на реальну фотографію (Decoder), показано на рисунку 1.13.

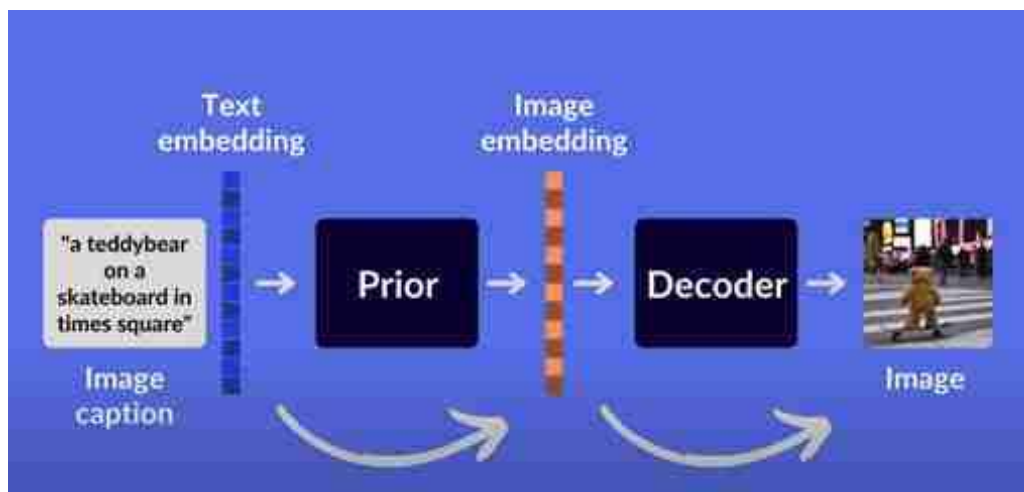


Рисунок 1.13 - Технологія DALL-E 2

Текст і зображення, що використовуються, надходять з іншої нейронної мережі під назвою CLIP (Contrastive Language-Image Pre-training), також створеної OpenAI. CLIP — це нейронна мережа, яка повертає найкращий підпис для вхідного зображення. Він робить протилежне до того, що робить DALL-E 2 – перетворює текст на зображення. Приклад цього процесу приведено на рис. 1.14.



Рисунок 1.14 – Приклад відтворюваних зображень

Метою DALL-E 2 є навчання двох моделей. Перший — Prior, навчений приймати текстові мітки та створювати вбудовані зображення CLIP. Другий — це

декодер, який використовує вбудовані зображення CLIP і створює навчене зображення. Після навчання робочий процес виведення виглядає наступним чином.

Введений підпис перетворюється на вбудовування CLIP-тексту за допомогою нейронної мережі.

Наступним кроком Prior зменшує розмірність вбудованого тексту за допомогою PCA — Principal Component Analysis (аналіз головних компонентів).

Вбудовування зображення створюється за допомогою вбудовування тексту.

На етапі декодування для перетворення використовується дифузійна модель («transformed image»). Після цього роздільна здатність кінцевого зображення збільшується з 64×64 до 256×256 , а потім до 1024×1024 за допомогою згорткової нейронної мережі [16, 17].

DALL-E 2 має аналогічну ситуацію як у Stable Diffusion, офіційна документація стосовно «prompt» відсутня, але також існує так звана «Prompt Book», в якій знаходиться більшість корисної інформації стосовно використання тексту. Книжку можна знайти за посиланням [18].

Щоб розпочати роботу з НМ необхідно виконати ряд послідовних кроків.

Для користувача взаємодія з нейронною мережею відбувається за допомогою веб інтерфейсу [19]. Його приклад показано на рисунку 1.15. Процес взаємодії користувача з мережею є найпростішою з розглянутих НМ, але для роботи також вимагає попередньої реєстрації облікового запису openAI.

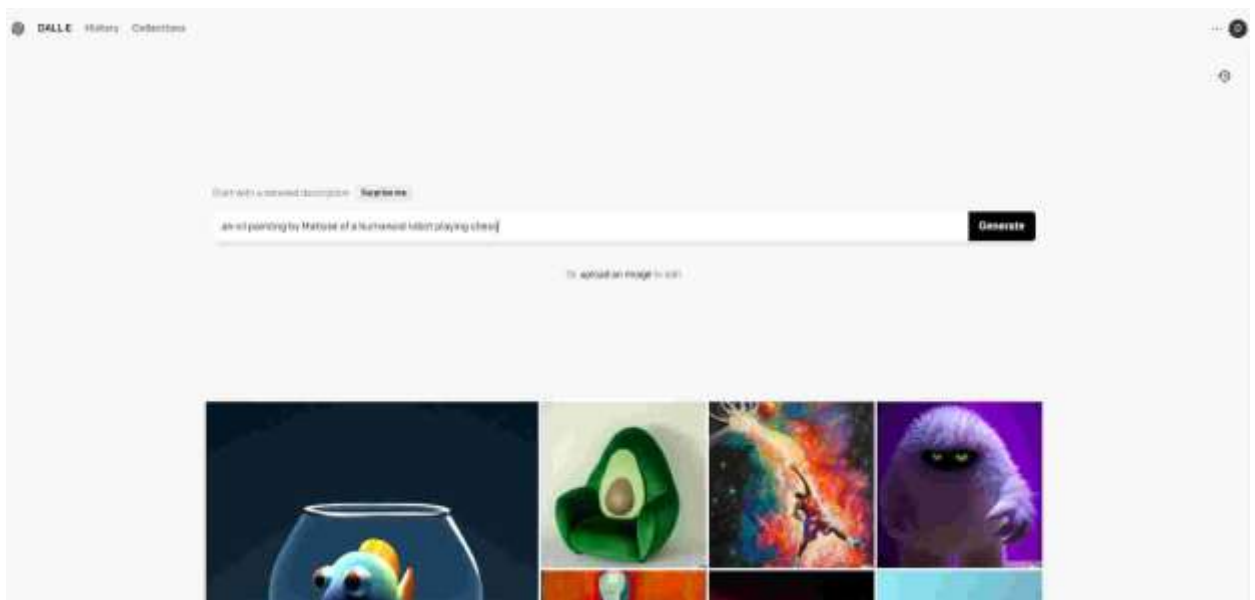


Рисунок 1.15- Веб інтерфейс DALL-E 2

DALL-E 2 поширюється на платній основі аналогічно Midjourney але за іншим методом, не залучаючи систему підписи, а так званими «credits», (рис. 1.16). Credits рівноцінні одній генерації зображення. Варто відзначити, що користувачам, які вже користувалися даною системою, раз на місяць надається 15 безкоштовних кредитів, які згорають по завершенню місяця.

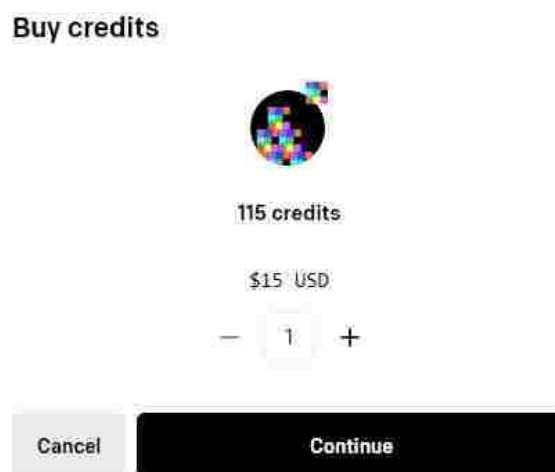


Рисунок 1.16- Ціна на купівлю кредитів

2 ГЕНЕРАЦІЯ ЗОБРАЖЕНЬ З ВИКОРИСТАННЯМ НЕЙРОННИХ МЕРЕЖ

2.1 Основні теоретичні положення генерації зображень

Робота НМ спирається на початкові дані, що подаються на її вхід (у даному разі це Prompt).

НМ може генерувати зображення, спираючись тільки на так звану текстову основу (prompt). Також є можливість виконувати дану роботу, використовуючи за основу вже існуюче зображення. У цьому разі НМ, аналізуючи вхідне зображення, бере з нього невказані основні необхідні параметри, наприклад, положення в просторі, кут камери, наближення, кольори та ін. В іншому випадку усі ці параметри будуть згенеровані випадковим чином і не завжди вони забезпечують задовільний результат.

Основні положення генерації Prompt:

Prompt - це так звані підказки НМ, які безпосередньо описують, що необхідно згенерувати.

Prompt engineering — це концепція штучного інтелекту, зокрема обробки природної мови. Інженерія «підказок» зазвичай працює шляхом перетворення одного або кількох завдань у набір даних на основі підказок і навчання мовної моделі за допомогою того, що називається «навчанням на основі підказок» або просто «навчанням підказками» [20].

Усі НМ, які проходять розбір, використовують дану систему, але програмний підхід у них відрізняється, тому і шанс отримання однакових зображень різними системами при використанні однакових «підказок» близький до нуля.

Також у prompt в деяких випадка задаються системні налаштування, на кшталт співвідношення сторін, необхідної роздільної здатності зображення, використання певної версії моделі та ін.

2.2 Складання початкової мети для аналізу та порівняння

НМ можуть підійти для безлічі задач, та без уточнення отриманий результат завжди буде відрізнятися. Тому для порівняння і аналізу сформовано точні теми для генерації зображень які наведені у вигляді таблиці з уточненим формулюванням, що саме необхідно отримати:

Таблиця 2.1 – Перелік тем генерації зображень

Назва теми	Запит
Відмінності у версіях	кіт
Будівля	житловий, багатоповерховий будинок
Людина	людина, обличчя, портрет
Пейзаж	природній, пейзаж міста, космічний простір, пейзаж у аніме стилістиці
Текстура	поверхня, абстракція, форми(фігури)
Автомобіль	міське авто
Листівка	листівка, обкладенка
Розгляд інших інструментів	Розгляд існуючих інструментів що пропонуються НМ

2.3 Робота із Midjourney

Взаємодія відбувається в Discord. Midjourney наразі доступний лише через бота Discord на офіційному сервері Discord, шляхом прямого обміну повідомленнями з ботом або запрошенням бота на сторонній сервер. Для створення зображення, користувачі використовують команду /imagine і вводять «prompts»; потім бот повертає набір із чотирьох зображень. Користувачі можуть вибрати, які

зображення вони хочуть збільшити. Midjourney також працює над веб-інтерфейсом.

При генерації зображень, сформульовано наступні запити:

2.3.1 Відмінності у версіях на прикладі кота

На рисунках 2.1 - 2.2 наведено зображення, згенеровані дотримуючись визначення наведеного у пункті 2.2.



Рисунок 2.1 – Зразок генерації кота версіями --v 5.1 та --niji 5

Опис використаних у рис. 2.1 «prompt»:

- ліва частина: «cat --v 5.1». «cat» означає «кіт», «--v 5.1» означає «version» тобто використану версію моделі 5.1;
- права частина: «cat --niji 5». «--niji 5» є використаною версією моделі niji.



Рисунок 2.2 – Зразок генерації кота версіями --v 3 та --v 2

Опис використаних у рис. 2.2 «prompt»:

ліва частина: «cat --v 3». «--v 3» означає використану версію моделі 3;

права частина: «cat --v 2». «--v 2» означає використану версію моделі 2.

Отримані результати відповідають встановленим вимогам.

2.3.2 Будівля

На рисунках 2.3 – 2.4 наведено зображення, згенеровані дотримуючись визначення наведеного у пункті 2.2.



Рисунок 2.3 – Зразок генерації житлового будинку

Опис використаних у рис. 2.3 «prompt»: «residential building» означає «житловий будинок».



Рисунок 2.4 – Зразок генерації високоповерхової будівлі

Опис використаних у рис. 2.4 «prompt»: «high-rise residential building in city -v 5.1». «high-rise residential building in city» означає «високоповерховий житловий будинок у місті».

Отримані результати відповідають встановленим вимогам.

2.3.3 Людина

На рисунках 2.5 – 2.8 наведено зображення, згенеровані дотримуючись визначення наведеного у пункті 2.2.



Рисунок 2.5 – Зразок генерації людини на вулиці

Опис використаних у рис. 2.5 «prompt»: «man, standing on city street». «man» означає «людина», «standing on city street» означає «стояча на міській вулиці».



Рисунок 2.6 – Зразок генерації обличчя та реалістичного обличчя

Опис використаних у рис. 2.6 «prompt»:

- ліва частина: «face». «face» означає «обличчя»;
- права частина: «realistic face --v 5.1». «realistic face», означає «реалістичне обличчя».



Рисунок 2.7 – Зразок генерації обличчя за допомогою --niji 5

Опис використаних у рис. 2.7 «prompt»:

- ліва частина: «realistic face --niji 5»;
- права частина: «face --niji 5».



Рисунок 2.8 – Зразок генерації портрету

Опис використаних у рис. 2.8 «prompt»: «50 years old man in black suit, white shirt, Rose in his lapel, bowtie, blue eyes, looking at viewer, portrait --ar 9:16 --v 5.1»:

- «50 years old man in black suit» - «50ти річний похилий чоловік у чорному костюмі»;

- «white shirt» - «біла сорочка»;

- «rose in his lapel» - «троянда в його відвороті»;

- «bowtie» - «краватка-метелик»;

- «blue eyes» - «блакитні очі»;

- «looking at viewer» - «дивиться на глядача»;

- «portrait» - «портрет»;

- «--ar 9:16» - «aspect ratio», тобто співвідношення сторін 9 до 16.

Отримані результати відповідають встановленим вимогам.

2.3.4 Пейзаж

На рисунках 2.9 – 2.12 наведено зображення, згенеровані дотримуючись визначення наведеного у пункті 2.2.



Рисунок 2.9 – Зразок генерації пейзажу

Опис використаних у рис. 2.9 «prompt»: «scenery», означає «пейзаж».



Рисунок 2.10 – Зразок генерації міського пейзажу та пейзажу сучасного міста

Опис використаних у рис. 2.10 «prompt»:

- ліва частина: «city scenery --v 5.1». «city scenery» означає «міський пейзаж»;
- права частина: «modern city scenery --v 5.1». «modern city scenery» означає «пейзаж сучасного міста».



Рисунок 2.11– Зразок генерації пейзажу

Опис використаних у рис. 2.11 «prompt»: «space scenery --v 5.1». «space scenery» означає «пейзаж».



Рисунок 2.12 – Зразок генерації аніме пейзажу--v 5.1 та --niji 5

Опис використаних у рис. 2.12 «prompt»:

- ліва частина: «anime scenery --v 5.1». «anime scenery» означає «аніме пейзаж» тобто пейзаж у стилістиці аніме;

- права частина: «anime scenery --niji 5».

Отримані результати відповідають встановленим вимогам.

2.3.5 Текстура

На рисунках 2.13 – 2.16 наведено зображення, згенеровані дотримуючись визначення наведеного у пункті 2.2.

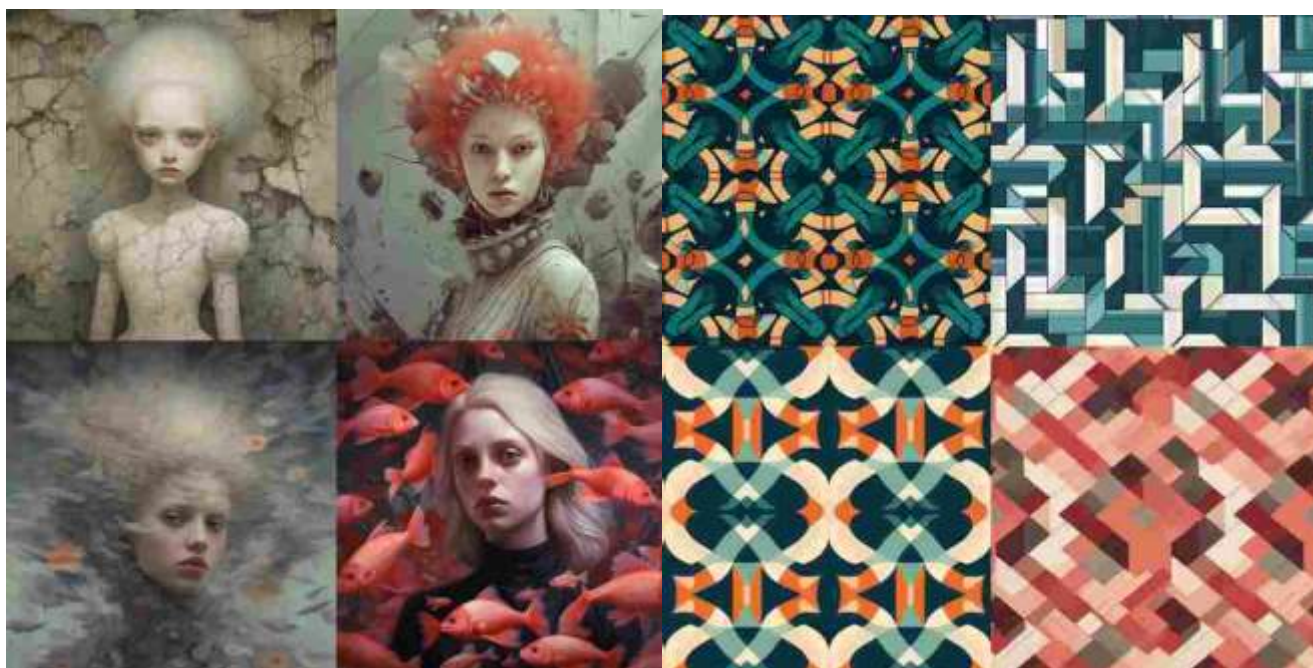


Рисунок 2.13 – Зразок генерації текстури за замовчуванням та з параметром --tile

Опис використаних у рис. 2.13 «prompt»:

- ліва частина: «texture --v 5.1». «texture» означає «текстура»;

- права частина: «texture --tile --v 5.1». «--tile» це параметр, що визначає задачу генерації зображення, які можна використовувати як повторювані плитку для створення безшовних візерунків.



Рисунок 2.14 – Зразок генерації текстури із запитом «структура каміння» та «реалістична структура каміння»

Опис використаних у рис. 2.14 «prompt»:

- ліва частина: «stone texture --tile --v 5.1». «stone texture» означає «текстура каміння»;
- права частина: «realistic stone texture --tile --v 5.1». «realistic stone texture» означає «реалістична текстура каміння».



Рисунок 2.15 – Зразок генерації абстрактної текстури за замовчуванням та з параметром --tile

Опис використаних у рис. 2.15 «prompt»:

- ліва частина: «abstraction --v 5.1». «abstraction» означає «абстракція»;
- права частина: «abstraction --tile --v 5.1».



Рисунок 2.16 – Зразок генерації текстури з використанням фігур за замовчуванням та з параметром --tile

Опис використаних у рис. 2.16 «prompt»:

- ліва частина: «texture shapes --v 5.1». «texture shapes» означає «текстура форми (фігури)»;
- права частина: «shapes --tile --v 5.1». «shapes» означає «форма (фігура)».

Отримані результати відповідають встановленим вимогам.

2.3.6 Автомобіль

На рисунках 2.17 – 2.18 наведено зображення, згенеровані дотримуючись визначення наведеного у пункті 2.2.



Рисунок 2.17 - Зразок генерації автомобіля за замовчуванням та реалістичного

Опис використаних у рис. 2.17 «prompt»:

- ліва частина: «car --v 5.1». «car» означає «автомобіль».
- права частина: «realistic car --v 5.1».



Рисунок 2.18 – Зразок генерації сучасного реалістичного автомобіля

Опис використаних у рис. 2.18 «prompt»: «realistic modern car --v 5.1». «modern» означає «сучасний».

Отримані результати відповідають встановленим вимогам.

2.3.7 Листівка

На рисунках 2.19 – 2.20 наведено зображення, згенеровані дотримуючись визначення наведеного у пункті 2.2.



Рисунок 2.19 – Зразок генерації листівки

Опис використаних у рис. 2.19 «prompt»: «flyer --v 5.1», означає «листівка».



Рисунок 2.20 – Зразок генерації обкладинки

Опис використаних у рис. 2.20 «prompt»: «the cover --v 5.1». «the cover» означає «обкладинка».

Отримані результати відповідають встановленим вимогам.

2.3.8 Розгляд інших інструментів

Інструмент аналізу зображення. Для використання необхідно задати команду /describe та завантажити зображення. Демонстрацію приведено на рис. 2.21 – 2.25.

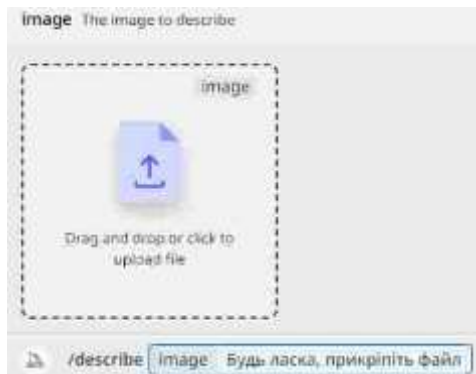


Рисунок 2. 21 – Завантаження файлу до НМ

Інструмент генерує чотири приклади підказок на основі зображення, яке було завантажуєте.



Рисунок 2.22 – Виведення результату опису зображення

Можливо також виконати генерацію схожих зображень на основі оригіналу (рис 2.23).



Рисунок 2.23 – Виведення схожих зображень з додатковим використанням «prompt» «dog»

Передбачена функція покращення роздільної здатності зображення (рис. 2.24, 2.25). Опис використаних у рис. 2.24 «prompt»: «a brown and white shiba inu dog sitting in the street, associated press photo --ar 58:75», означає «коричнево-біла собака шіба-іну сидить на вулиці, пов'язане фото для преси», --ar 58:75 – співвідношення сторін 58 до 75.



Рисунок 2.24 – Початкова генерація зображень з використанням «prompt»



Рисунок 2.25 – Отримане зображення у високій роздільній здатності

Можливо також виконати об'єднання декількох зображень в єдине, демонстрацію приведено на рис. 2.26 – 2.27



Рисунок 2.26 – Зразки оригіналу зображення

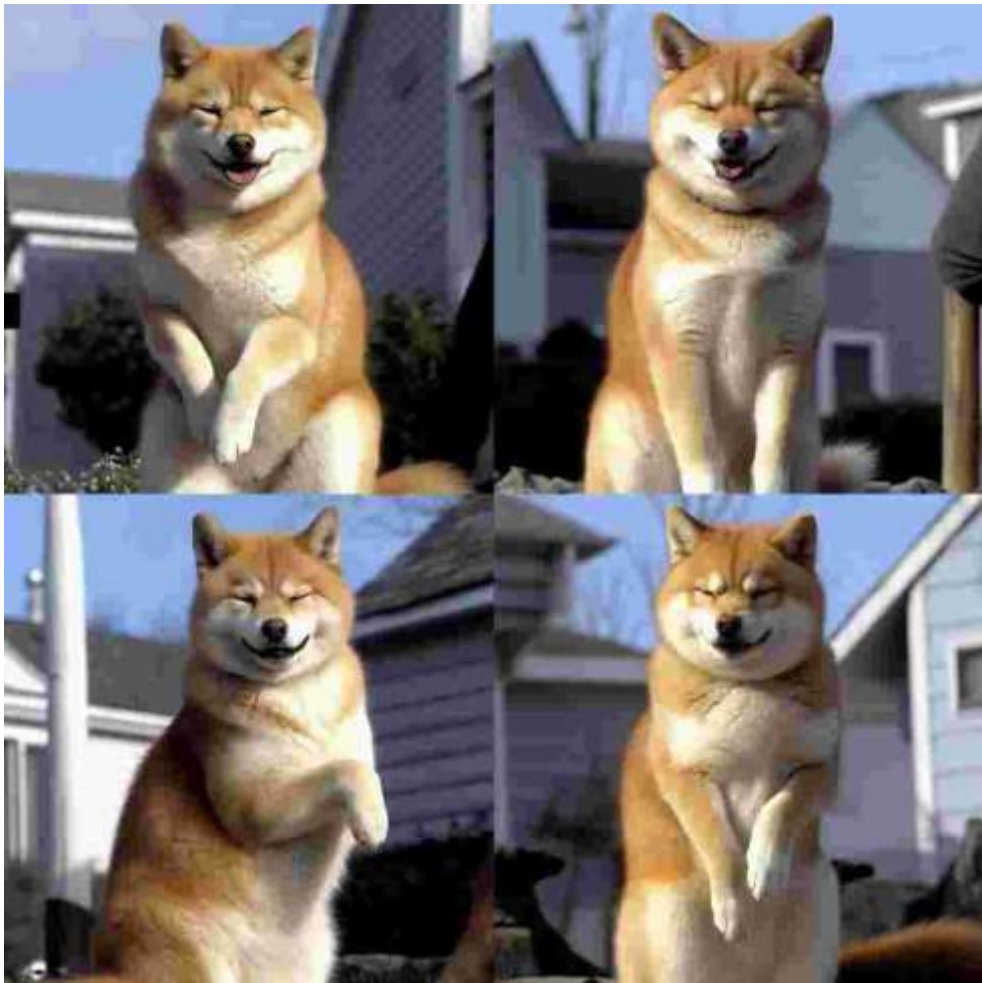


Рисунок 2.27 – Похідні версії від об'єднання оригіналів

Іструмент об'єднання зображень Midjourney є унікальним з розібраних НМ та при цьому надає гарні результати, що є перевагою серед інших розглянутих НМ.

2.4 Робота із Stable Diffusion

Взаємодія відбувається або через онлайн сервіси, або встановлюючи собі програмне забезпечення. Використовуючи онлайн сервіси втрачається одна з переваг у гнучкості налаштувань, з іншого боку для першого знайомства це може бути зручніше.

У даному дипломному проекті робота зі Stable Diffusion виконана із використанням персонального комп'ютеру в якості робочої станції, характеристики якої подані у таблиці 2.2.

Таблиця 2.2 – таблиця основних компонентів ПК для керівників

Комплектуючі	Найменування
Процесор	AMD Ryzen 5 3600X
Оперативна пам'ять	DDR4 8Gb x2 3200 МГц
Відеокарта	GeForce® RTX 2080 SUPER
SSD-накопичувач	m2 1 ТБ
Операційна система	Windows 10

У випадку використання на власному комп'ютері взаємодія відбувається через веб інтерфейс.

Для зручності та наочності подальшого порівняння і аналізу при роботі з даною НМ буде задано параметр Batch count «4» для генерації набору з чотирьох зображень.

Опис параметрів, використаних у процесі генерації, наведено у таблиці 2.3.

Таблиця 2.3 - Опис параметрів при генерації Stable Diffusion

Параметр	Опис
Steps	Sampling Steps, кількість ітерацій. Чим більше, тим довше генерується і більше деталей на зображенні.
Sampler	Sampling Method, математична функція, що визначає правила обробки вхідного шуму. Це впливає на зображення та швидкість генерації.
CFG Scale	(classifier-free guidance scale) – це величина відповідності текстовому запиту
Seed	Зерно для випадкової генерації
Size	Роздільна здатність зображення
Model hash	Контрольна сума використаної моделі
Model	Використана модель
ENSD	Початкова дельта шуму Eta
(:)	Підвищення «ваги» параметру
Denoising strength	Коефіцієнт зниження шуму

При генерації зображень, використано запити, наведені в таблиці 2.1.

2.4.1 Відмінності у версіях на прикладі кота

На рисунках 2.28 – 2.30 наведено зображення, згенеровані дотримуючись визначення, наведеного у пункті 2.2.



Рисунок 2.28 – Зразок генерації кота

Використані налаштування та «prompt»: cat Steps: 20, Sampler: Euler a, CFG scale: 7, Seed: 300210063, Size: 512x512, Model hash: e1441589a6, Model: SD , ENSD: 31337

Результат наведено на рис. 2.28.



Рисунок 2.29 – Зразок генерації кота з використанням моделі SD 512-base-ema

Використані налаштування: cat Steps: 20, Sampler: Euler a, CFG scale: 7, Seed: 1953942008, Size: 512x512, Model hash: ab800444cd, Model: SD 512-base-ema, ENSD: 31337

Результат наведено на рис. 2.29.



Рисунок 2.30 – Зразок генерації кота з використанням моделі SD 768-v-ema

Використані налаштування: cat Steps: 30, Sampler: Euler a, CFG scale: 7, Seed: 189838462, Size: 512x512, Model hash: bfcaf07557, Model: SD 768-v-ema, ENSD: 31337

Результат наведено на рис. 2.30.

Отримані результати відповідають встановленим вимогам.

2.4.2 Будівля

На рисунках 2.31 – 2.33 наведено зображення, згенеровані дотримуючись визначення наведеного у пункті 2.2.



Рисунок 2.31 – Зразок генерації будівлі

Використані налаштування: residential building Steps: 20, Sampler: Euler a, CFG scale: 7, Seed: 1564841075, Size: 512x512, Model hash: e1441589a6, Model: SD, ENSD: 31337

Результат наведено на рис. 2.31.

Отримане зображення вийшло не достатньо чітким, тому було додано додаткові «prompt», а саме «best quality, extremely detailed, masterpiece» які означають «найкраща якість, надзвичайно деталізований, шедевр». Також були використані «negative prompt», а саме «sketches, (worst quality:2), (low quality:2), (normal quality:2), lowres, normal quality, cropped, jpegartifacts, signature, watermark, username, ((blurry))» які означають «ескізи, (найгірша якість: 2), (низька якість: 2), (нормальна якість: 2), низька роздільна здатність, нормальна якість, обрізаний, артефакти картинки формату jpeg, підпис, водяний знак, ім'я користувача, ((розмитість))»



Рисунок 2.32 – Зразок генерації будівлі з уточнюючим запитом

Використані налаштування: residential building, best quality, extremely detailed, masterpiece, best quality, extremely detailed Negative prompt: sketches, (worst quality:2), (low quality:2), (normal quality:2), lowres, normal quality, cropped, jpegartifacts, signature, watermark, username, ((blurry)) Steps: 20, Sampler: Euler a, CFG scale: 7, Seed: 1564841075, Size: 512x512, Model hash: e1441589a6, Model: model, ENSD: 31337

Результат наведено на рис. 2.32.

Зображення вийшло більш чітким але отриманий результат змінився, бо у вхідних даних було вказано просто житловий будинок, без типу будівлі (котедж, хмарочос та ін), тому «prompt» «residential building» замінено на «high-rise residential building» приведено на рис. 2.33.



Рисунок 2.33 – Зразок генерації багатоповерхової будівлі з уточнюючим запитом

Використані налаштування: high-rise residential building, best quality, extremely detailed, masterpiece, best quality, extremely detailed Negative prompt: sketches, (worst quality:2), (low quality:2), (normal quality:2), lowres, normal quality, cropped, jpegartifacts, signature, (watermark), username, ((blurry)) Steps: 20, Sampler: Euler a, CFG scale: 7, Seed: 1564841075, Size: 512x512, Model hash: e1441589a6, Model: model, ENSD: 31337

Результат наведено на рис. 2.33.

Отримані результати відповідають встановленим вимогам.

2.4.3 Людина

На рисунках 2.34 – 2.38 наведено зображення, згенеровані дотримуючись визначення, наведеного у пункті 2.2.



Рисунок 2.34 – Зразок генерації людини

Використані налаштування: man, standing on city street Steps: 20, Sampler: Euler a, CFG scale: 7, Seed: 986340376, Size: 512x512, Model hash: e1441589a6, Model: SD, ENSD: 31337

Результат наведено на рис. 2.34.



Рисунок 2.35 – Зразок генерації обличчя

Використані налаштування: face Steps: 20, Sampler: Euler a, CFG scale: 7, Seed: 1564841075, Size: 512x512, Model hash: e1441589a6, Model: SD, ENSD: 31337

Результат наведено на рис. 2.35.

Отриманий результат вийшов художнім, для запобігання цьому додано «prompt» «realistic»

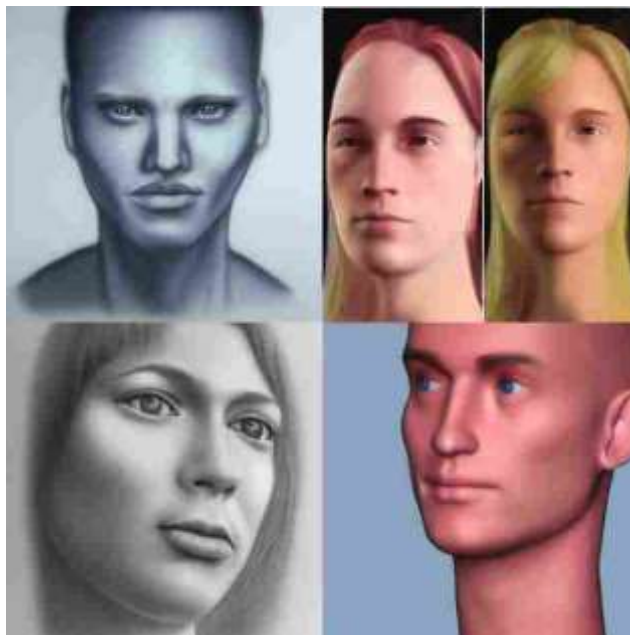


Рисунок 2.36 – Зразок генерації реалістичного обличчя

Використані налаштування: realistic face Steps: 20, Sampler: Euler a, CFG scale: 7, Seed: 29816399, Size: 512x512, Model hash: e1441589a6, Model: SD, ENSD: 31337

Результат наведено на рис. 2.36.

Результат все ще не прийнятний, в наступну ітерацію додано «highly detailed real», які означають «високодеталізоване реальне»

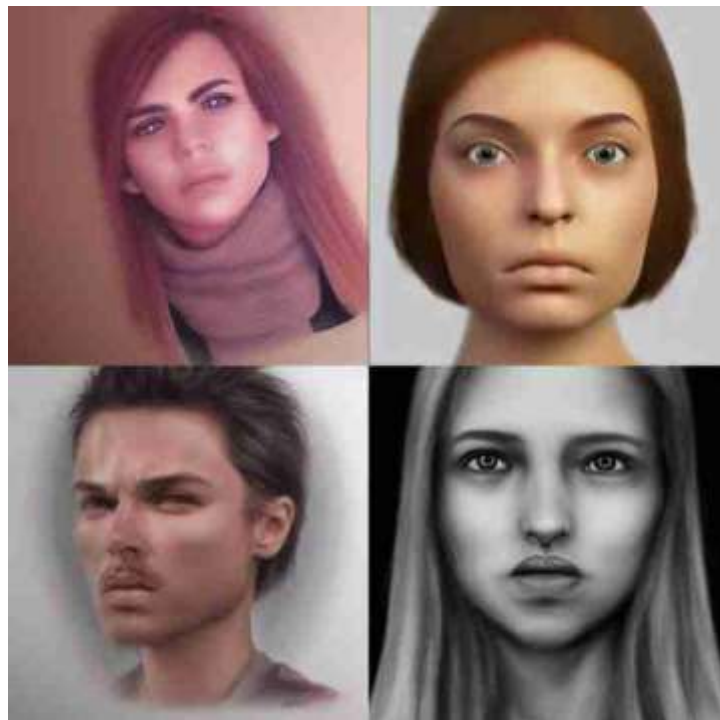


Рисунок 2.37 – Зразок генерації високодеталізованого реального обличчя

Використані налаштування: realistic highly detailed real face Steps: 20, Sampler: Euler a, CFG scale: 7, Seed: 3967936850, Size: 512x512, Model hash: e1441589a6, Model: SD, ENSD: 31337

Результат наведено на рис. 2.37.



Рисунок 2.38 – Зразок генерації портрету

Використані налаштування: Sitting on chair, in dark room, Black_suit,Black_jacket,shirt,tuxedo,red_Rose in his lapel, (Close-Up:1.0),(face_focus:1.0),(upper_body), pubic_hair, bowtie, blue_eyes, brown watch, 1 old_Man, 50yo,male,looking at viewer,(closed_mouth:1.0),official art,extremely detailed CG unity 8k wallpaper, perfect lighting,Colorful, Bright_Front_face_Lighting, (masterpiece:1.0),(best_quality:1.0), ultra high res,4K,ultra-detailed, photography, 8K, HDR, highres, absurdres:1.2, Kodak portra 400, film grain, blurry background, bokeh:1.2, lens flare, (vibrant_color:1.2) Negative prompt: EasyNegative, bad-artist-anime, bad-hands-5, lowres, bad anatomy, bad hands, text, error, missing fingers, extra digit, fewer digits, cropped, jpeg artifacts, signature, watermark, username, blurry, out of focus, censorship, Blurry faces, Blank faces, bad face, Ugly, extra ear, amputee, missing hands, missing arms, missing legs, Extra fingers, 6 fingers, Extra feet, ghost, Extra legs, Extra hands, (painting by bad-artist-anime:0.9), (painting by bad-artist:0.9), text, error, blurry, jpeg artifacts, cropped, normal quality, artist name, (worst quality, low quality:1.4),twisted_hands,fused_fingers, Steps: 27, Sampler: Euler a, CFG scale: 7, Seed: 586633887, Size: 576x928, Model hash: bfcaf07557 and e1441589a6, Model: SD 768-v-ema and SD, ENSD: 31337

Результат наведено на рис. 2.38.

Отримані результати відповідають встановленим вимогам.

2.4.4 Пейзаж

На рисунках 2.39 – 2.42 наведено зображення, згенеровані дотримуючись визначення, наведеного у пункті 2.2.



Рисунок 2.39 – Зразок генерації пейзажу

Використані налаштування: scenery Steps: 20, Sampler: Euler a, CFG scale: 7, Seed: 1564841075, Size: 512x512, Model hash: e1441589a6, Model: SD, ENSD: 31337
Результат наведено на рис. 2.39.



Рисунок 2.40 – Зразок генерації міського пейзажу

Використані налаштування: city scenery Steps: 30, Sampler: Euler a, CFG scale: 7, Seed: 1866319013, Size: 512x512, Model hash: e1441589a6, Model: SD, ENSD: 31337

Результат наведено на рис. 2.40.



Рисунок 2.41 – Зразок генерації космічного пейзажу

Використані налаштування: space scenery Steps: 30, Sampler: Euler a, CFG scale: 7, Seed: 1181771193, Size: 512x512, Model hash: e1441589a6, Model: SD, ENSD: 31337

Результат наведено на рис. 2.41.



Рисунок 2.42 – Зразок генерації пейзажу в стилі аніме

Додано додатковий приклад у іншій популярній на сьогоднішній день стилістиці, для демонстрації можливостей. Використані налаштування: anime scenery Steps: 20, Sampler: Euler a, CFG scale: 6, Seed: 1864144866, Size: 512x512, Model hash: e1441589a6, Model: SD, ENSD: 31337

Результат наведено на рис. 2.42.

Отримані результати відповідають встановленим вимогам.

2.4.5 Текстура

На рисунках 2.43 – 2.45 наведено зображення, згенеровані дотримуючись визначення, наведеного у пункті 2.2.

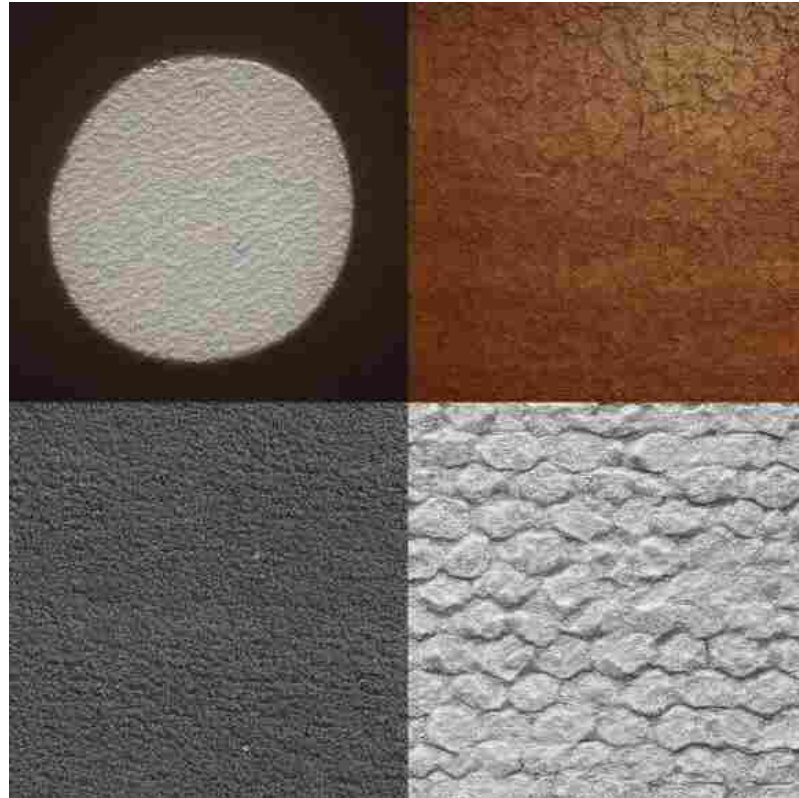


Рисунок 2.43 – Зразок генерації текстури

Використані налаштування: texture Steps: 20, Sampler: Euler a, CFG scale: 7, Seed: 3260306112, Size: 512x512, Model hash: e1441589a6, Model: SD, ENSD: 31337

Результат наведено на рис. 2.43.



Рисунок 2.44 – Зразок генерації абстрактної текстури

Використані налаштування: abstraction Steps: 30, Sampler: Euler a, CFG scale: 7, Seed: 248068505, Size: 512x512, Model hash: e1441589a6, Model: SD, ENSD: 31337

Результат наведено на рис. 2.44.



Рисунок 2.45 – Зразок генерації текстури з фігур

Використані налаштування: shapes Steps: 30, Sampler: Euler a, CFG scale: 7, Seed: 31210976, Size: 512x512, Model hash: e1441589a6, Model: SD, ENSD: 31337

Результат наведено на рис. 2.45.

Отримані результати відповідають встановленим вимогам.

2.4.6 Автомобіль

На рисунку 2.46 наведено зображення, згенеровані дотримуючись визначення, наведеного у пункті 2.2.



Рисунок 2.46 – Зразок генерації автомобіля

Використані налаштування: car Steps: 20, Sampler: Euler a, CFG scale: 7, Seed: 848936560, Size: 512x512, Model hash: e1441589a6, Model: SD, ENSD: 31337

Результат наведено на рис. 2.46.

Отримані результати відповідають встановленим вимогам.

2.4.7 Листівка

На рисунках 2.47 – 2.48 наведено зображення, згенеровані дотримуючись визначення, наведеного у пункті 2.2.



Рисунок 2.47 – Зразок генерації листівки

Використані налаштування: flyer Steps: 20, Sampler: Euler a, CFG scale: 7, Seed: 2243451029, Size: 512x512, Model hash: e1441589a6, Model: SD, ENSD: 31337
Результат наведено на рис. 2.47.



Рисунок 2.49 – Похідні версії від оригіналу зображення

Використані налаштування:

Dog Steps: 20, Sampler: Euler a, CFG scale: 7, Seed: 499367353, Size: 512x512, Model hash: e1441589a6, Model: SD, Denoising strength: 0.75, ENSD: 31337

Результат наведено на рис. 2.49.

Підвищення роздільної здатності існуючого зображення наведено на плакаті ПЛЗ. Даний інструмент Stable Diffusion, в свою чергу, є унікальним серед розглянутих НМ, та дозволяє підвищувати роздільну здатність будь яких зображень. Цей інструмент може стати у нагоді навіть поза генерацією зображень.

2.5 Робота із DALL-E 2

Взаємодія відбувається через веб сайт за посиланням [18], шляхом вводу даних у веб інтерфейс.

При генерації зображень, використано запити, наведені в таблиці 2.1.

2.5.1 Відмінності у версіях на прикладі kota

На рисунку 2.50 наведено зображення, згенеровані дотримуючись визначення, наведеного у пункті 2.2.



Рисунок 2.50 – Зразок генерації kota

Опис використаних у рис. 2.50 «prompt»: «cat».

2.5.2 Будівля

На рисунках 2.51 – 2.52 наведено зображення, згенеровані дотримуючись визначення, наведеного у пункті 2.2.



Рисунок 2.51 – Зразок генерації житлової будівлі

Опис використаних у рис. 2.51 «prompt»:
«Residential building».



Рисунок 2.52 – Зразок генерації багатоповерхової житлової будівлі

Опис використаних у рис. 2.52 «prompt»:
«High-rise residential building».

Отримані результати відповідають встановленим вимогам.

2.5.3 Людина

На рисунках 2.53 – 2.56 наведено зображення, згенеровані дотримуючись визначення, наведеного у пункті 2.2.



Рисунок 2.53 – Зразок генерації людини на вулиці міста

Опис використаних у рис. 2.53 «prompt»:
«man, standing on city street».



Рисунок 2.54 – Зразок генерації реалістичного обличчя

Опис використаних у рис. 2.54 «prompt»:

«Realistic face».



Рисунок 2.55 – Зразок генерації обличчя як малюнок олівцем

Опис використаних у рис. 2.55 «prompt»:

«Face, pencil drawing». «pencil drawing» означає «малюнок олівцем».



Рисунок 2.56 – Зразок генерації портрету

Опис використаних у рис. 2.56 «prompt»:

«50 years old man in black suit, white shirt, Rose in his lapel, bowtie, blue eyes, looking at viewer, portrait».

Отримані результати відповідають встановленим вимогам.

2.5.4 Пейзаж

На рисунках 2.57 – 2.60 наведено зображення, згенеровані дотримуючись визначення наведеного у пункті 2.2.



Рисунок 2.57 – Зразок генерації пейзажу

Опис використаних у рис. 2.57 «prompt»: «scenery».



Рисунок 2.58 – Зразок генерації міського пейзажу

Опис використаних у рис. 2.58 «prompt»: «city scenery».



Рисунок 2.59 – Зразок генерації космічного пейзажу

Опис використаних у рис. 2.59 «prompt»: «space scenery».



Рисунок 2.60 – Зразок генерації пейзажу у стилі аніме

Опис використаних у рис. 2.60 «prompt»:

«anime scenery».

Отримані результати відповідають встановленим вимогам.

2.5.5 Текстура

На рисунках 2.61 – 2.63 наведено зображення, згенеровані дотримуючись визначення, наведеного у пункті 2.2.



Рисунок 2.61 – Зразок генерації текстури

Опис використаних у рис. 2.61 «prompt»:

«Texture».



Рисунок 2.62 – Зразок генерації абстрактної текстури

Опис використаних у рис. 2.62 «prompt»: «abstraction».



Рисунок 2.63 – Зразок генерації текстури з використанням фігур

Опис використаних у рис. 2.63 «prompt»: «a shapes».

Отримані результати відповідають встановленим вимогам.

2.5.6 Автомобіль

На рисунках 2.64 – 2.66 наведено зображення, згенеровані дотримуючись визначення, наведеного у пункті 2.2.



Рисунок 2.64 – Зразок генерації автомобіля

Опис використаних у рис. 2.64 «prompt»:
«car».



Рисунок 2.65 – Зразок генерації реалістичного автомобіля

Опис використаних у рис. 2.65 «prompt»: «Realistic car».



Рисунок 2.66 – Зразок генерації автомобіля при їзді по дорозі

Опис використаних у рис. 2.66 «prompt»:
«car driving on the road». «driving on the road» означає «їзда по дорозі».
Отримані результати відповідають встановленим вимогам.

2.5.7 Листівка

На рисунках 2.67 – 2.71 наведено зображення, згенеровані дотримуючись визначення, наведеного у пункті 2.2.



Рисунок 2.67 – Зразок генерації листівки

Опис використаних у рис. 2.67 «prompt»: «flyer».

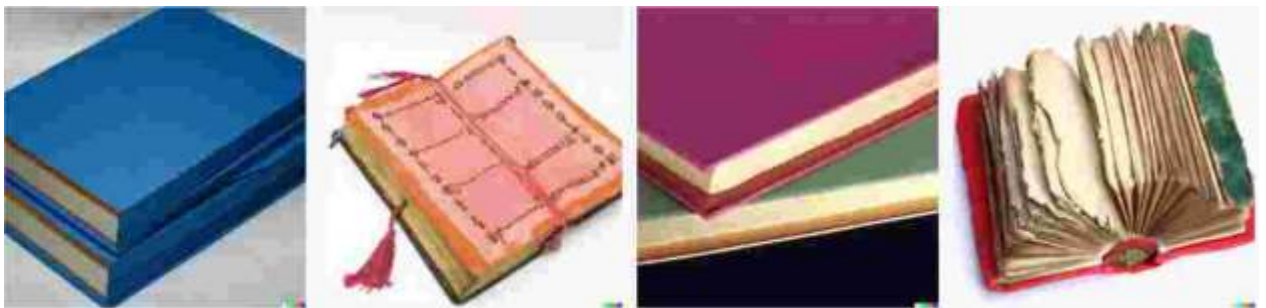


Рисунок 2.68 – Зразок генерації листівки як обкладинки

Опис використаних у рис. 2.68 «prompt»: «The cover».

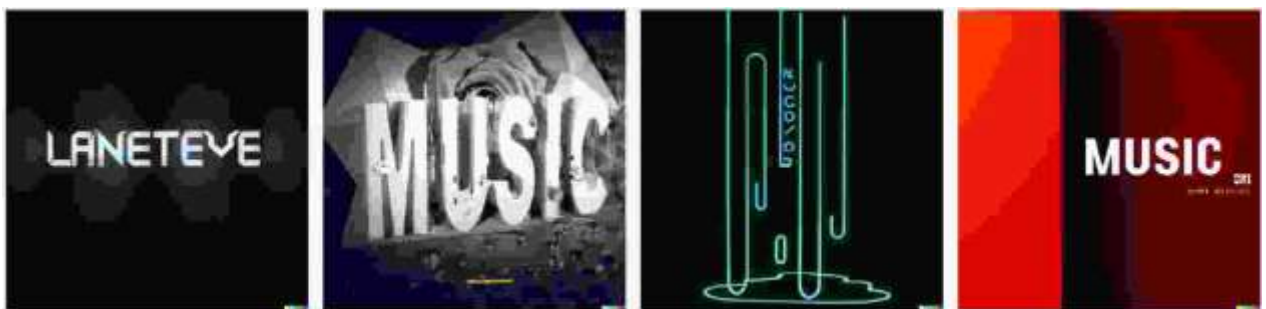


Рисунок 2.69 – Зразок генерації листівки як обкладинки музичного альбому

Опис використаних у рис. 2.69 «prompt»:

«Music album cover». «Music album cover» означає «Обкладинка музичного альбому».

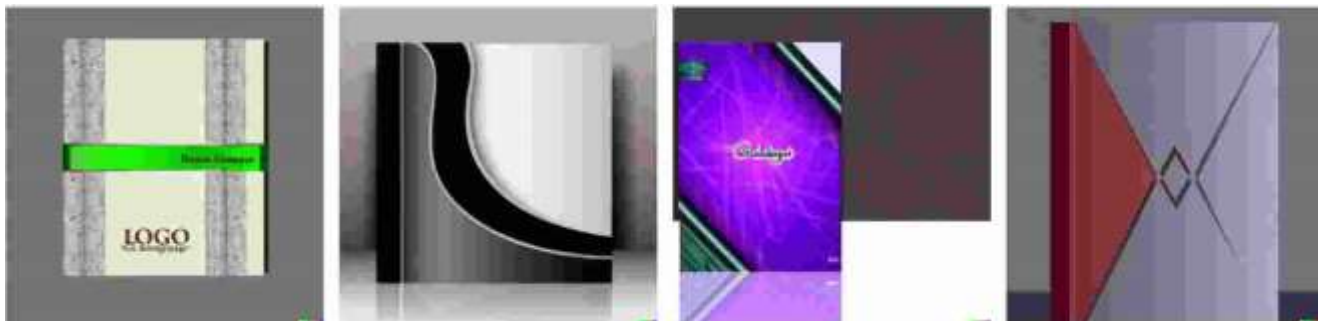


Рисунок 2.70 – Зразок генерації листівки як обкладинки книги

Опис використаних у рис. 2.70 «prompt»:

«Book cover design». «Book cover design» означає «Дизайн обкладинки книги».

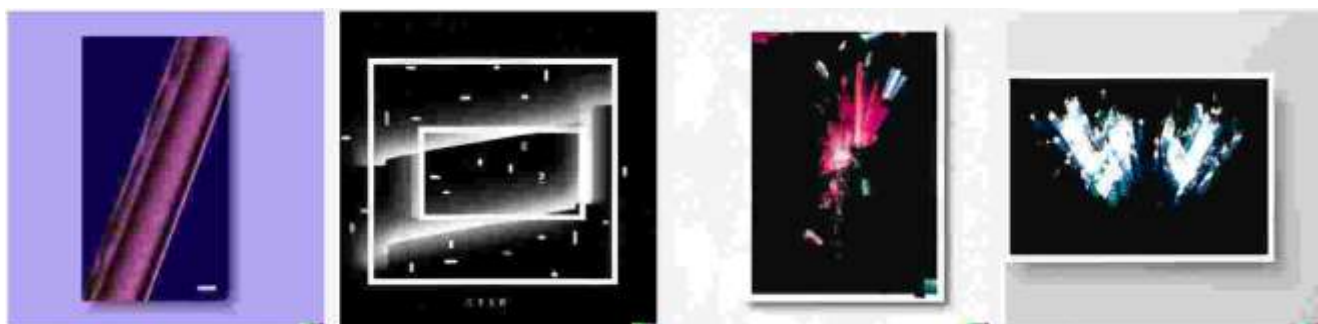


Рисунок 2.71 – Зразок генерації листівки як обкладинки альбому
в абстрактному стилі

Опис використаних у рис. 2.71 «prompt»:

«Album cover in abstract style». «Album cover in abstract style» означає «Обкладинка альбому в абстрактному стилі».

Отримані результати відповідають встановленим вимогам.

2.5.8 Розгляд інших інструментів

Додатковим інструментом є можливість генерації схожих зображень на основі оригіналу, реалізацію приведено на рис. 2.72 – 2.73.



Рисунок 2.72 – Завантаження оригіналу зображення



Рисунок 2.73 – Похідні версії від оригіналу зображення

Результат редагування зображення приведено на рис. 2.74 – 2.78.



Рисунок 2.74 – Обрізане зображення



Рисунок 2.75 – Використані «prompt» «Dog sitting on city street»



Рисунок 2.76 – Використані «prompt» «Dog playing Frisbee»



Рисунок 2.77 – Використані «prompt» «Dog sitting on the porch»



Рисунок 2.78 – Використані «prompt» «Cat hugging with dog»

DALL-E 2 краще генерує похідні від оригіналу зображення серед інших розглянутих НМ, використовуючи при цьому меншу кількість уточнень. Також DALL-E 2 має унікальний серед розглянутих НМ інструмент інтерактивного редагування зображення, та при цьому має «user-friendly» інтерфейс.

3 ПОРІВНЯННЯ ТА АНАЛІЗ РОБОТИ НМ ДЛЯ ГЕНЕРАЦІЇ ЗОБРАЖЕНЬ

Отримавши результати генерації можна проаналізувати та порівняти отримані результати.

3.1 Відмінності у версіях

У Midjourney, та Stable Diffusion присутній доступ до різних версій, на відміну від DALL-E 2, в якій можна користуватися лише найостаннішою.



Рисунок 3.1 – Приклад порівняння Midjourney v2 та v5.1

При розвитку версій явно помітно що НМ стає більш детально «розміти» об'єкт що генерується.



Рисунок 3.2 – Приклад порівняння Stable Diffusion v1.5 та v2

При розвитку версій НМ стає більш складною та при переході на різні версії необхідні додаткові налаштування.



Рисунок 3.3 – Приклад генерації кота останньою версією DALL-E 2

3.2 Робота із Midjourney

Midjourney чудово справляється зі створенням естетично привабливих мальовничих зображень із взаємодоповнюючими кольорами, художнім використанням світла й тіні та задовільною композицією. Відомий своєю послідовністю. Незважаючи на те, що створені зображення не зовсім фотореалістичні, вони досить близькі, та при бажанні їх можна використовувати для редагування у різних дизайнерських програмах.



Рисунок 3.4 – Midjourney

Midjourney також працює через інтерфейс на основі Discord, який може бути корисним для тих, хто не знайомий з Discord.

Зрештою, Midjourney може похвалитися однією з найбільших спільнот штучного інтелекту з 14,5 мільйонами учасників Discord, які надсилають від 20 до 40 завдань щосекунди.

3.3 Робота із Stable Diffusion

Stable Diffusion може генерувати детальні зображення на основі текстових описів. Має відкритий вихідний код, та може працювати на будь яких ПК із графічним процесором із принаймні 6 ГБ відеопам'яті. Він дає результати, які можна порівняти з DALL-E 2 і MidJourney, але йому бракує «user-friendly» інтерфейсу користувача, але це проблема тільки для недосвідченого користувача бо присутня можливість переробити інтерфейс під себе при достатньому рівні знань. Користувачі мають права на створені зображення, хоча модель викликала суперечки щодо етики власності та впливу на людей-митців. Stable Diffusion є більш доступним у створенні вмісту порівняно з іншими продуктами на основі ШІ.



Рисунок 3.5 – Stable Diffusion

У даній НМ для отримання задовільного результату необхідно досить точно задавати чином доведеться дуже довго перебирати «prompt» для задовільного результату, наступні кроки пропущені.

Перша ітерація з великою вірогідністю не буде відповідати визначеним вимогам. При подальших ітераціях для отримання задовільного результату необхідно додати додаткові «prompt» уточнюючи необхідний результат.

3.4 Робота із DALL-E 2

DALL·E 2 - це передова система штучного інтелекту, призначена для створення зображень із текстових описів, яка пропонує безліч функцій, які дозволяють покращувати та редагувати зображення. Здатний створювати майже фотореалістичні зображення з роздільною здатністю, що вчетверо кращі, ніж його попередня версія DALL·E. Цей потужний інструмент також може імітувати різні художні стилі та маніпулювати існуючими зображеннями для створення нових композицій або варіацій.



Рисунок 3.6 – DALL·E 2

Хоча DALL·E 2 має вражаючі можливості, це все ще технологія, що розвивається, хоча це і стосується галузі в цілому.

3.5 Рекомендації та методика використання НМ.

Для отримання реалістичного зображення рекомендовано користуватися Stable Diffusion або DALL-E 2. Основним стилем даних НМ є реалістичні зображення. Використовуючи «prompt» наведені у таблиці 3.1 результат буде ще більш близьким до бажаного

Таблиця 3.1 – Опис «prompt»

«prompt»	Опис підказки
realistic	реалістичний
realistic style	реалістичний стиль
photography	фотографія
digital painting	цифровий живопис
use camera	використовуючи камеру
photorealism	фотореалізм

Для отримання графічного ефекту зображення рекомендовано користуватися Stable Diffusion, дана НМ найкраще підійде для отримання картинок, які виглядають ніби 3D графіка. Рекомендовані «prompt» для даної задачі приведені у таблиці 3.2.

Таблиця 3.2 – Опис «prompt»

«prompt»	Опис підказки
3D	3D, тривимірний простір
render	візуалізація, відтворення
unreal engine	зроблений на ігровому 3д рушію
unity	зроблений на ігровому 3д рушію
CG	Комп'ютерна графіка

Для отримання зображення у популярній в останній час стилістиці аніме рекомендовано скористатися Midjourney. У даній НМ присутня додаткова модель «niji» яка була спеціально створена для генерації зображень у аніме стилістиці. Для генерації зображень у Midjourney за допомогою моделі «niji» слід використовувати

«prompt» «--niji», який явно вказує НМ що необхідно використовувати саме дану модель.

Також зображення у аніме стилістиці можна отримати за допомогою Stable Diffusion, але для цього треба завантажити користувацькі неофіційні моделі, знайти які можна за посиланнями [15, 22].

Для генерації схожих зображень, подібних оригіналу, рекомендовано використовувати DALL-E 2. Skorистavшись DALL-E 2, це можна зробити найбільш зручно, та отримавши результат не гірше інших НМ, а в деяких випадках навіть краще, наприклад без використання додаткових уточнюючих підказок.

Для підвищення роздільної здатності можна скористатись лише інструментами Stable Diffusion, через те що інші розглянуті НМ не мають такого функціоналу. Midjourney має можливість збільшення роздільної здатності зображення, але лише згенерованих самим Midjourney, та лише до 1024x1024 пікселів, Stable Diffusion в свою чергу не має таких обмежень, та може працювати із будь якими зображеннями.

Таким чином, для генерації певного зображення доцільно виконати наступні кроки:

а) в залежності від бажаної стилістики зображення, що очікується, а також типу вхідних даних, обрати оптимальний інструмент, спираючись на запропоновані рекомендації.

б) задати вхідні дані для опису очікуваного зображення:

1) задати ключове слово зображення або завантажити базове зображення;

2) задати при необхідності уточнюючі запити;

в) згенерувати зображення;

г) якщо отримане зображення не відповідає запиту, повторити пункти 2 та 3 до отримання необхідного рівня якості.

Дана послідовність дій є універсальною для отримання зображення, спираючись на словесний опис та на базове зображення з додатковим описом.

ВИСНОВКИ

Дипломний проект присвячений дослідженню якості генерації зображень різних стилістик з використанням нейронних мереж.

В ході виконання дипломного проекту виконано наступне:

- розглянуто основні положення щодо принципів роботи нейронних мереж для генерації зображень;
- виконано огляд різних типів нейронних мереж та аналіз інструментів генерації зображень за їх допомогою;
- описано процес роботи з розібраними типами нейронних мереж;
- проведено порівняння та аналіз роботи нейронних мереж стосовно якості результатів генерації зображень;
- запропоновано методику генерації зображень різної стилістики.

Так, Midjourney та DALL-E 2 мають нижчий поріг входження для початку роботи з ними через те, що навіть використавши декілька слів виходить задовільний результат.

В Stable Diffusion зазвичай при першій ітерації без додаткових уточнень зображення виходить лише віддалено схоже на запитуване. При роботі з цією нейронною мережею доводиться задавати більше параметрів, ніж при роботі з іншими нейронними мережами. Так, параметр для усунення дефектів зображення в Stable Diffusion необхідно задавати явно, на відміну від Midjourney та DALL-E 2, де він враховується автоматично та використовується лише для видалення зайвих речей. Проте дана НМ серед розглянутих є безкоштовною, та надає можливість використання користувацьких версій її розширення та надбудов.

При роботі було витрачено 45 доларів на придбання доступу до сервісів НМ Midjourney та DALL-E 2.

Результати аналізу згенерованих зображень показали, що для отримання задовільної картинки з мінімальною кількістю уточнюючих запитів доцільніше використовувати Midjourney, для отримання графічного зображення перевагу

краще надати Stable Diffusion, для стилю аніме – Midjourney(niji) або Stable Diffusion використовуючи користувальницькі моделі. При цьому задача генерації нового зображення за вже існуючим краще вирішується за допомогою DALL-E 2.

Таким чином, галузь генерації зображень за допомогою нейронних мережами має безліч варіацій використання та функціональних можливостей, перелік яких з часом розширюється та вдосконалюються, що дозволяє використовувати даний напрямок для вирішення різних дизайнерських задач.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

- 1 Стаття про “foundation models” розвиток III URL: <https://www.economist.com/interactive/briefing/2022/06/11/huge-foundation-models-are-turbo-charging-ai-progress> (дата звернення: 01.03.2023)
- 2 Стаття про сучасну генерацію зображень використовуючи HM URL: <https://theconversation.com/give-this-ai-a-few-words-of-description-and-it-produces-a-stunning-image-but-is-it-art-184363> (дата звернення: 01.03.2023)
- 3 OpenAI дебютує з DALL-E для створення зображень із тексту URL: <https://venturebeat.com/business/openai-debuts-dall-e-for-generating-images-from-text/> (дата звернення: 12.03.2023)
- 4 Офіційний ресурс продукту DALL-E 2 від OpenAI URL: <https://openai.com/product/dall-e-2> (дата звернення: 08.03.2023)
- 5 Оголошення про запуск Stable Diffusion URL: <https://stability.ai/blog/stable-diffusion-announcement> (дата звернення: 09.03.2023)
- 6 Запис інтерв'ю взятого у Emad Mostaque, засновника Stability AI URL: <https://www.youtube.com/watch?v=YQ2QtKcK2dA> (дата звернення: 10.03.2023)
- 6 Стаття про початок роботи Firefly Adobe зі штучним інтелектом URL: <https://techcrunch.com/2023/03/21/adobe-firefly-generative-ai/> (дата звернення: 08.03.2023)
- 7 Офіційна новина Adobe про представлення Firefly, сімейство нового Creative Generative AI URL: <https://news.adobe.com/news/news-details/2023/Adobe-Unveils-Firefly-a-Family-of-new-Creative-Generative-AI/default.aspx> (дата звернення: 03.03.2023)
- 8 Стаття огляд продукту Craiyon URL: <https://golden.com/wiki/Craiyon-ZY8KR8P> (дата звернення: 05.03.2023)
- 9 Midjourney Documentation URL: <https://docs.midjourney.com/> (дата звернення: 13.04.2023)
- 10 Midjourney URL: <https://www.midjourney.com/> (дата звернення: 08.03.2023)

11 CLIP: Connecting text and images URL: <https://openai.com/research/clip> (дата звернення: 07.03.2023)

12 Stable Diffusion prompts search engine URL: <https://stablediffusionweb.com/prompts> (дата звернення: 13.04.2023)

13 Stable Diffusion Prompt Book URL: <https://openart.ai/promptbook> (дата звернення: 13.04.2023)

14 WebUI Stable Diffusion розміщена на GitHub URL: <https://github.com/AUTOMATIC1111/stable-diffusion-webui> (дата звернення: 15.03.2023)

15 Transformers library Hugging Face URL: <https://huggingface.co/> (дата звернення: 15.03.2023)

16 Стаття про роботу DALL-E 2, Stable Diffusion і Midjourney URL: <https://www.marktechpost.com/2022/11/14/how-do-dall%C2%B7e-2-stable-diffusion-and-midjourney-work/> (дата звернення: 09.03.2023)

17 Згортова нейронна мережа – просте пояснення CNN та її застосування URL: <https://evergreens.com.ua/ua/articles/cnn.html> (дата звернення: 05.03.2023)

18 The DALL·E 2 Prompt Book URL: <https://dallery.gallery/the-dalle-2-prompt-book/> (дата звернення: 21.04.2023)

19 НМ генерації зображень DALL-E 2 URL: <https://labs.openai.com/> (дата звернення: 08.03.2023)

20 Дослідницька робота Language Models are Unsupervised Multitask Learners URL: https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf (дата звернення: 07.03.2023)

21 Midjourney Statistics approachableAI URL: <https://approachableai.com/midjourney-statistics/#:~:text=Midjourney's%20Discord%20has%2014.5%20million,dramatic%20traffic%20growth%20in%202023.> (дата звернення: 13.05.2023)

22 CivitAI URL: <https://civitai.com/> (дата звернення: 19.04.2023)

Порівняння згенерованих зображень різних НМ



Рисунок 1 - приклад генерації Stable Diffusion



Рисунок 2 - приклад генерації Midjourney



Рисунок 3 - приклад генерації DALL-E 2

					13.02070849.00038 ПЛ1					
					Розробка методики генерації зображень з використанням нейронних мереж	Лист.		Маса	Масшт.	
Зам.	Лист.	№ докум.	Підп.	Дата	Порівняння згенерованих зображень різних НМ					
Розроб.		Семельянов Д.В.								
Перевір.		Голуб Т.В.								
Т.контр.										
						Лист 1		Листів 1		
Н.контр.		Щербак Н.В.				НУ "Запорізька політехніка"				
Затвердж.		Кудерметов Р.К.				КНТ 510 - СП				

Додаткові інструменти НМ Midjourney



Рисунок 4 - Генерація зображень використовуючи інструмент опису НМ



Рисунок 5 - Підвищення роздільної здатності обраного зображення



Рисунок 2 - Похідні версії від оригіналу зображення



Рисунок 3 - Підвищення роздільної здатності обраного зображення

- 1 a brown and white shiba inu dog sitting in the street, associated press photo --ar 58:75
- 2 a white and brown dog sits on the sidewalk, in the style of japanese influence, light orange and dark crimson, associated press photo, strong facial expression, smilecore, northern and southern dynasties, sharp/prickly --ar 58:75
- 3 a dog is smiling while sitting on a brick wall, in the style of japanese influence, associated press photo, ken sugimori, shuzo oshimi, red and amber, animal intensity --ar 58:75
- 4 a dog sitting next to some tall grass, in the style of japanese, light maroon and white, strong facial expression, tooth wu, light orange and silver, associated press photo, sanriocore --ar 58:75



Рисунок 1 - Генерація опису зображення

					13.02070849.00038 ПЛ2				
					Розробка методики генерації зображень з використанням нейронних мереж Додаткові інструменти НМ Midjourney	Лист.		Маса	Масшт.
Зам.	Лист.	№ докум.	Підп.	Дата					
Розроб.	Семельянов Д.В.								
Перевір.	Голуб Т.В.								
Т.контр.						Лист 1		Листів 1	
						НУ "Запорізька політехніка" КНТ 510 - СП			
Н.контр.	Щербак Н.В.								
Затвердж.	Кудерметов Р.К.								

Додаткові інструменти НМ Stable Diffusion



Рисунок 1 - Оригінал зображення



Рисунок 2 - Підвищення роздільної здатності зображення



Рисунок 3 - Похідні версії від оригіналу зображення

					13.02070849.00038 ПЛЗ					
					Розробка методики генерації зображень з використанням нейронних мереж	Лит.		Маса	Масшт.	
Зам.	Лист.	№ докум.	Підп.	Дата	Додаткові інструменти НМ Stable Diffusion	Лист 1		Листів 1		
Розроб.		Семельянов Д.В.								
Перевір.		Голуб Т.В.								
Т.контр.										
Н.контр.		Щербак Н.В.								
Затвердж.		Кудерметов Р.К.								
						НУ "Запорізька політехніка"				
						КНТ 510 - СП				

Додаткові інструменти НМ DALL·E 2



Рисунок 2 – Генерація фону зображення



Рисунок 3 – Редагування фону зображення



Рисунок 4 - Редагування зображення



Рисунок 5 – Об'єднання декількох зображень



Рисунок 1 - Оригінал зображення

					13.02070849.00038 ПЛ4				
					Розробка методики генерації зображень з використанням нейронних мереж Додаткові інструменти НМ DALL·E 2	Лит.		Маса	Масшт.
Зам.	Лист.	№ докум.	Підп.	Дата					
Розроб.		Семельянов Д.В.							
Перевір.		Голуб Т.В.							
Т.контр.									
						Лист 1		Листів 1	
Н.контр.		Щербак Н.В.				НУ "Запорізька політехніка"			
Затвердж.		Кудерметов Р.К.				КНТ 510 - СП			