

**ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ
ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ
КАФЕДРА КОМП'ЮТЕРНИХ НАУК**

КВАЛІФІКАЦІЙНА РОБОТА
на тему: «Використання штучного інтелекту для генерації
зображень»

на здобуття освітнього ступеня магістра
зі спеціальності 122 Комп'ютерні науки

(код, найменування спеціальності)

освітньо-професійної програми Комп'ютерні науки

(назва)

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

(підпис)

Сергій КУЧМА

(Ім'я, ПРІЗВИЩЕ здобувача)

Виконав:
здобувач вищої освіти
група КНДМ-62

Сергій КУЧМА

Керівник:

*науковий ступінь,
вчене звання*

Ірина ЩЕРБИНА

к.т.н., доцент

Рецензент:

*науковий ступінь,
вчене звання*

(Ім'я, ПРІЗВИЩЕ)

Київ 2023

**ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ**
Навчально-науковий інститут інформаційних технологій

Кафедра Комп'ютерних наук

Ступінь вищої освіти Магістр

Спеціальність Комп'ютерні науки

Освітньо-професійна програма Комп'ютерні науки

ЗАТВЕРДЖУЮ

Завідувач кафедри Комп'ютерних наук

_____ Віктор ВИШНІВСЬКИЙ
« _____ » _____ 2023 р.

**ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ**

_____ Кучмі Сергію Володимировичу

(прізвище, ім'я, по батькові здобувача)

1. Тема кваліфікаційної роботи: Використання штучного інтелекту для генерації зображень

керівник кваліфікаційної роботи Ірина ЩЕРБИНА к.т.н., доцент,

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

затверджені наказом Державного університету інформаційно-комунікаційних технологій від «19» 10.2023р. №145

2. Строк подання кваліфікаційної роботи «29» грудня 2023р.

3. Вихідні дані до кваліфікаційної роботи: Технології генерації зображень за допомогою штучного інтелекту, науково-технічна література

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

Дослідження методів та інструментів для генерації зображень

Вибір технічних засобів для створення зображень

Розробка та експерименти з генерацією зображень на основі вибраних алгоритмів

5. Перелік графічного матеріалу: *презентація*

1. Розвиток ШІ, методи генерації

2. Алгоритми та моделі штучного інтелекту
3. Проведення експериментів та збір результатів
4. Виявлення сильних та слабких сторін розробленої моделі

6. Дата видачі завдання «19» жовтня 2023 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів кваліфікаційної роботи	Строк виконання етапів роботи	Примітка
1	Аналіз наявної науково-технічної літератури	19.10-05.11.23	
2	Дослідження поставленої задачі	05.11-12.11.23	
3	Вивчення документації програмного забезпечення	13.11-19.11.23	
4	Розробка імплементації моделі	20.11-25.11.23	
5	Проведення експериментів та збір результатів	27.11-03.12.23	
6	Аналіз отриманих результатів	04.12-10.12.23	
7	Оформлення роботи: вступ, висновки, реферат	11.12-20.12.23	
8	Розробка демонстраційних матеріалів	21.12-29.12.23	

Здобувач вищої освіти

(підпис)

Сергій КУЧМА

(Ім'я, ПРІЗВИЩЕ)

Керівник
кваліфікаційної роботи

(підпис)

Ірина ЩЕРБИНА

(Ім'я, ПРІЗВИЩЕ)

РЕФЕРАТ

Текстова частина кваліфікаційної роботи на здобуття освітнього ступеня магістра:
60 стор., 45 рис., 18 джерел.

Мета роботи – дослідження штучного інтелекту та промптів для створення зображень.

Об'єкт дослідження – різновиди штучного інтелекту для створення та редагування зображень.

Предмет дослідження – програмне забезпечення, промпти та згенеровані зображення.

Короткий зміст роботи: програмне забезпечення штучного інтелекту, генерація й редагування зображень та їхнє використання. В роботі проведено аналіз генерації зображень штучним інтелектом, розглянуто алгоритми створення зображень та сучасні технічні засоби, промпти. В практичній частині виконане написання промптів, створення зображень та їхнє порівняння.

КЛЮЧОВІ СЛОВА: зображення, штучний інтелект, промпт, генерація, ефективність, інноваційність, ШІ.

ABSTRACT

Text part of the master's qualification work:

60 pages, 45 pictures, 18 sources.

The purpose of the work – research on artificial intelligence and image-processing software.

Object of research – types of artificial intelligence for creating and editing images.

Subject of research – software, prompts and generated images.

Summary of the work: artificial intelligence software, image generation and editing, and their use. The paper analyzes the generation of images by artificial intelligence, considers image creation algorithms and modern technical tools, prompts. In the practical part, the author writes prompts, creates images, and compares them.

KEYWORDS: image, artificial intelligence, prompt, generation, efficiency, innovation, AI.

ЗМІСТ

ВСТУП.....	9
1 ДОСЛІДЖЕННЯ ТЕХНОЛОГІЇ ГЕНЕРАЦІЇ ЗОБРАЖЕНЬ В ІНДУСТРІЇ ШТУЧНОГО ІНТЕЛЕКТУ	10
1.1 Місце технології генерації зображень в індустрії штучного інтелекту.....	10
1.2 Проблематика галюцинацій штучного інтелекту в генерації зображень.....	26
1.3 Дослідження етичних проблем, пов'язаних з використанням штучного інтелекту.....	34
2 ДОСЛІДЖЕННЯ ТЕХНОЛОГІЧНИХ РЕСУРСІВ ГЕНЕРАЦІЇ ЗОБРАЖЕНЬ.....	39
2.1 Сучасне програмне забезпечення для генерації зображень.....	39
2.2 Порівняння функціональних можливостей програмного забезпечення для генерації зображень.....	43
2.3 Методи виявлення штучно згенерованих зображень.....	53
2.4 Висновки.....	55
3 ПРАКТИЧНА РЕАЛІЗАЦІЯ.....	56
3.1 Дослідження та генерація зображень в Midjourney.....	56
3.2 Дослідження та генерація зображень в Dall-E.....	60
3.3 Дослідження та генерація зображень в DreamStudio.....	65
ВИСНОВКИ.....	71
ПЕРЕЛІК ПОСИЛАНЬ.....	72
ДОДАТОК 1.....	74

ВСТУП

Штучний інтелект (ШІ) і його застосування в різних галузях науки та техніки стали ключовими факторами для реалізації інноваційних та передових технологій. Однією з таких важливих галузей застосування ШІ є генерація зображень, яка є актуальною областю досліджень у світі комп'ютерної науки та штучного інтелекту.

Генерація зображень за допомогою штучного інтелекту є важливим напрямком, який відкриває нові можливості у сферах візуального мистецтва, дизайну, медіа та науки. Розвиток таких технологій дозволяє створювати реалістичні, абстрактні або фантастичні зображення, що знаходять застосування в різних галузях, від відтворення віртуальних світів до розробки унікальних художніх творів.

У даній магістерській роботі вивчається використання штучного інтелекту для генерації зображень, спрямоване на подальший розвиток та вдосконалення цих технологій. Метою дослідження є розробка та впровадження ефективної моделі для генерації зображень, що враховує сучасні досягнення в області штучного інтелекту та глибокого навчання.

1 ІСТОРІЯ СТВОРЕННЯ ТА РОЗВИТОК ШТУЧНОГО ІНТЕЛЕКТУ

1.1 Визначення та класифікація штучного інтелекту

Штучний інтелект (ШІ) є фундаментальною галуззю інформаційних технологій, що спрямована на створення систем, здатних виконувати задачі, які зазвичай вимагають людського інтелекту. Ця область науки та техніки прагне відтворити когнітивні функції людини, такі як розпізнавання образів, розуміння природної мови, прийняття рішень, та навіть творчість. [1]

Визначення ШІ включає в себе створення імітації інтелекту через використання алгоритмів, які здатні аналізувати дані, вчитися з досвіду, та робити розсудливі висновки. Ця область тісно пов'язана із концепцією машинного навчання, де алгоритми можуть самостійно вдосконалювати свою ефективність на основі накопиченого досвіду. [2]

Штучний інтелект можна класифікувати за кількома ключовими аспектами, враховуючи його завдання та функціональність:

- за інтелектом;
- за методом досягнення завдань;
- за областю застосування.

За інтелектом розрізняють сильний штучний інтелект та слабкий штучний інтелект.

Сильний ШІ: має за мету створити інтелект, еквівалентний або перевищуючий рівень людського інтелекту, включно з усіма аспектами когнітивної функції.

Слабкий ШІ: спрямований на вирішення конкретних завдань та обмежений в області взаємодії та аналізу, не має свідомості чи загального розуміння.

За методом досягнення завдань розрізняють символічний штучний інтелект та відданий даним штучний інтелект.

Символічний ШІ: використовує символи та правила для вирішення завдань, моделюючи людські здібності до логічного мислення та розуміння.

Відданий даним ШІ: орієнтований на роботу з великими обсягами даних, використовуючи алгоритми машинного навчання для виявлення паттернів та здійснення висновків.

За областю застосування розрізняють експертні системи, системи розпізнавання образів, природній мовний обробник та генерацію зображень.

Експертні системи: використовують знання фахівців для розв'язання конкретних завдань.

Системи розпізнавання образів: спрямовані на розпізнавання та аналіз зображень або відео.

Природній мовний обробник: аналізують та розуміють природну мову, дозволяючи взаємодію з користувачем.

Генерація зображень: аналізують запити та генерують зображення, користуючись заданими параметрами.

Класифікація штучного інтелекту є важливим аспектом розуміння цієї галузі, оскільки вона дозволяє розкрити різноманітність його застосувань та методів, що забезпечують реалізацію інтелектуальних систем та розвиток нових технологій.

Класифікація штучного інтелекту представлена на рисунку 1.1.

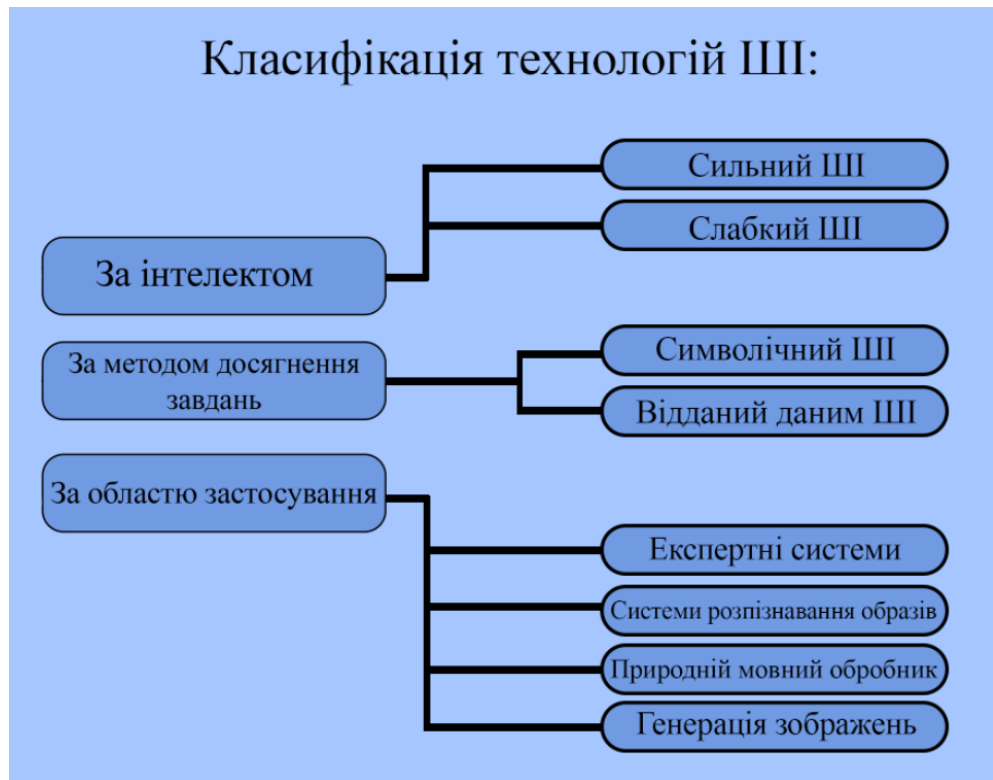


Рисунок 1.1 - Класифікація технологій ШІ

Визначимо основні технології використання штучного інтелекту [14][15][16][17][18].

Machine learning (Машинне навчання) – підмножина штучного інтелекту, яка передбачає навчання алгоритмів для прогнозування або прийняття рішень на основі вхідних даних. На відміну від традиційного програмування, де правила й логіка чітко визначені, алгоритми ML навчаються з використанням великих наборів даних для вивчення шаблонів і прийняття рішень.

Машинне навчання - це, по суті, форма прикладної статистики з підвищеним акцентом на використанні комп'ютерів для статистичного оцінювання складних функцій і зниженим акцентом на доведенні довірчих інтервалів навколо цих функцій.

Зазвичай науковців цікавить, наскільки добре алгоритм машинного навчання працює на даних, яких він раніше не бачив, оскільки це визначає, наскільки добре він буде працювати в реальному світі. Тому нейронну мережу оцінюють за

показниками ефективності, використовуючи тестовий набір даних, який відрізняється від даних, що використовуються для навчання системи машинного навчання.

В реальному світі моделі машинного навчання зазвичай використовуються для розпізнавання зображень, розпізнавання мовлення, виявлення шахрайства, персоналізації та охорони здоров'я.

Deep learning (Глибоке навчання) - це підмножина машинного навчання, яка передбачає навчання нейронних мереж із кількома рівнями для розпізнавання шаблонів у даних. Моделі глибокого навчання зазвичай використовуються для складних завдань, які включають великі обсяги даних, як-от розпізнавання зображень і мови, обробка природної мови та автономне водіння. [8][14]

Моделі глибокого навчання складаються з шарів штучних нейронів, кожен з яких обробляє вхідні дані та передає їх на наступний рівень. Ці рівні дозволяють моделі вивчати дедалі складніші варіації вхідних даних, зрештою роблячи прогнози або приймаючи рішення на основі відомих шаблонів.

Багато досліджень у галузі глибокого навчання пов'язані з побудовою ймовірнісної моделі вхідних даних, $p_{\text{model}}(x)$. Така модель може, в принципі, використовувати ймовірнісний висновок для передбачення будь-якої змінної в її оточенні, враховуючи будь-які інші змінні.

Лінійна факторна модель описує процес генерації даних наступним чином. Спочатку ми виберемо пояснювальні фактори h з розподілу $h \sim p(h)$, де $p(h)$ - факторний розподіл, з $p(h) = \prod_i p(h_i)$, а з цього випливає наступна схема (рис. 1.2):

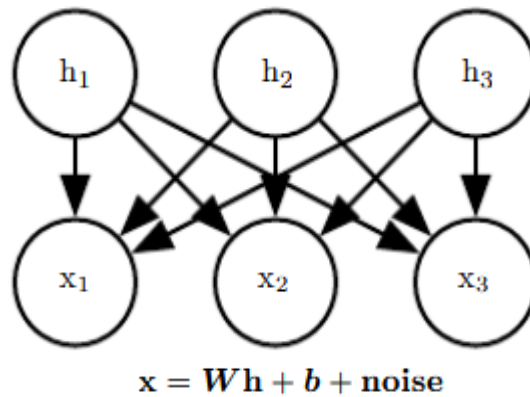


Рисунок 1.2 - Графічна модель, що описує сімейство лінійних факторних моделей

Natural language processing (Обробка природної мови) - це підгалузь штучного інтелекту (ШІ), яка передбачає навчання машин розумінню, інтерпретації та створенню людської мови. Вона дозволяє машинам обробляти та аналізувати великі обсяги текстових і мовних даних, створюючи такі програми, як переклад мови, аналіз настроїв і чат-боти.

Машинний переклад мови у його нинішньому вигляді зосереджений на вирішенні низки дуже практичних завдань. Мабуть, найпоширенішим застосуванням машинного перекладу в даний час є вільний доступ до інформації.

У світі існує близько 7000 мов. Деякі аспекти людської мови здаються універсальними, працюючими для кожної з цих мов. Але у кожній мові є свої особливості, не притаманні ні для якої іншої мови. На схемі представлено проблематику перекладу тексту з англійської мови на німецьку (рис. 1.3).



Рисунок 1.3 - Складність перекладу тексту з англійської на німецьку

Robotics (Робототехніка) - це галузь дослідження та інженерії, яка займається проектуванням, конструкцією та використанням роботів. Робот - це машина або автономний агент, здатний автоматично виконувати складну серію дій, як правило, за допомогою програмування або контролю за допомогою комп'ютера. Робототехніку використовують в промисловому виробництві, дослідженнях, сільському господарстві, військових роботах тощо. [15]

Одним з аспектів орієнтування робота у просторі є візуальне сприйняття зображень. Ми розглянемо лише перший крок у цьому складному процесі - формування ознак.

Робот зчитує зображення і генерує одну або кілька ознак для заданого зображення. Елементами зазвичай є скаляри (наприклад, площа або співвідношення сторін) або короткі вектори (наприклад, координати об'єкта або параметри лінії). Виділення елементів зображення є необхідним першим кроком у використанні даних зображення для керування роботом. Це крок концентрації інформації, який зменшує швидкість передачі даних на 10^6 - 10^8 байт на виході камери, щоб створити до порядку десятків ознак на кожен кадр, які можуть бути використані як вхідні дані для системи керування роботом (рис. 1.4).

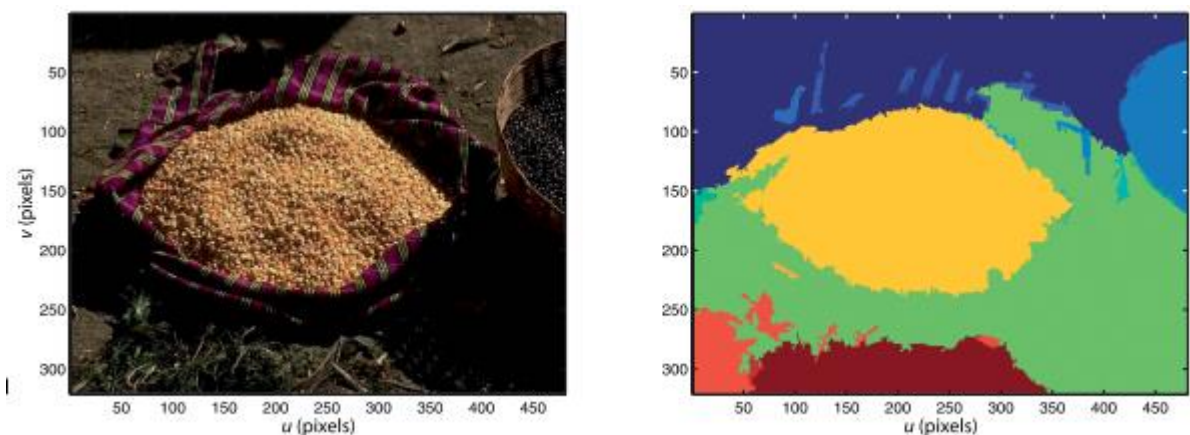


Рисунок 1.4 - Один із кроків штучного інтелекту створення ознак для заданого зображення

Computer vision (Комп'ютерний зір) - це галузь дослідження та інженерії, яка зосереджена на тому, щоб дозволити машинам інтерпретувати та аналізувати візуальні дані з навколишнього світу. Комп'ютерний зір передбачає розробку алгоритмів і методів, які дозволяють пристроям розпізнавати, обробляти та розуміти зображення та відео.

На практиці комп'ютерний зір включає в себе розпізнавання об'єктів, сегментацію зображень, оптичне розпізнавання символів віртуальну реальність та робототехніку.

Чому робота з комп'ютерним зором особливо складна? Частково тому, що це обернена задача, в якій ми намагаємося відновити деякі невідомі, маючи недостатньо інформації для повного визначення рішення. Тому ми мусимо вдаватися до фізичних та імовірнісних моделей або машинного навчання на великих наборах прикладів, щоб великих наборів прикладів, щоб розрізнити потенційні рішення.

Цікаво те, що люди і тварини вміють робити це без особливих зусиль, тоді як алгоритми комп'ютерного зору роблять банальні помилки.

Але є і випадки, в яких комп'ютерний зір має нескінченні переваги, як ось наприклад так звані “оптичні ілюзії” (рис. 1.5).

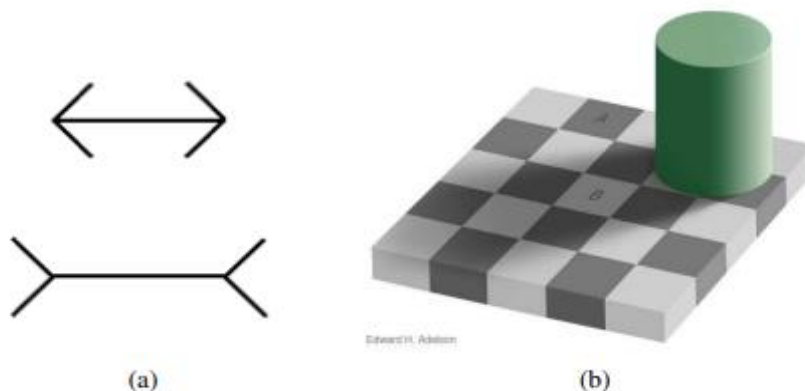


Рисунок 1.5 - Оптичні ілюзії; (a) Відрізки, які насправді мають однакову довжину. Клітинки А та В насправді однакового кольору (b)

Для людського ока здається, що відрізки мають різну довжину, а клітинки А та В абсолютно різного кольору. Але якщо підійти до аналізу картинки “цифровим” методом, як це робить комп’ютерний зір, то він не стане жертвою оптичної ілюзії.

Probability and statistics (Ймовірність і статистика) є важливою основою для розуміння штучного інтелекту. Будь-яка система штучного інтелекту, якій потрібно приймати рішення на основі даних, а потім оновлюватися на основі цих рішень, значною мірою покладається на обидві концепції розподілу ймовірностей.

Розподіл ймовірностей - це опис того, наскільки ймовірно, що випадкова величина або набір випадкових величин набуде кожного з можливих станів. Спосіб опису розподілу ймовірностей залежить від того, чи є змінні дискретними або неперервними.

Ймовірність надає можливість кількісної оцінки невизначеності, тоді як статистика дозволяє аналізувати дані та робити з них висновки – навіть дані, які традиційно не піддаються аналізу, наприклад слова, зображення чи звуки.

Neural Networks (Нейронні мережі) належать до алгоритмів машинного навчання та засновані на функціонуванні нейронів у мозку людини. Вони засновані на тому, що за певними параметрами існує спосіб їх поєднання для отримання певного результату. Дані проходять через різні рівні, на яких застосовуються ряди правил навчання, поки не досягнуть останнього рівня, де результати порівнюються з «правильними», а параметри коригуються на основі функції «ваги», заданої в кожному правилі. Після того, як мережа навчилася, вона може виправити свої «ваги» та працювати в режимі пам’яті або виконання.

Наприклад, перед тим як прийняти рішення, мозок людини один за одним аналізує десятки та сотні умов, які стосуються необхідного рішення (рис. 1.6). Схожим чином працюють і нейронні мережі. В якомусь сенсі “перебираючи” можливі варіанти, нейронна мережа вчиться приймати рішення краще і краще.

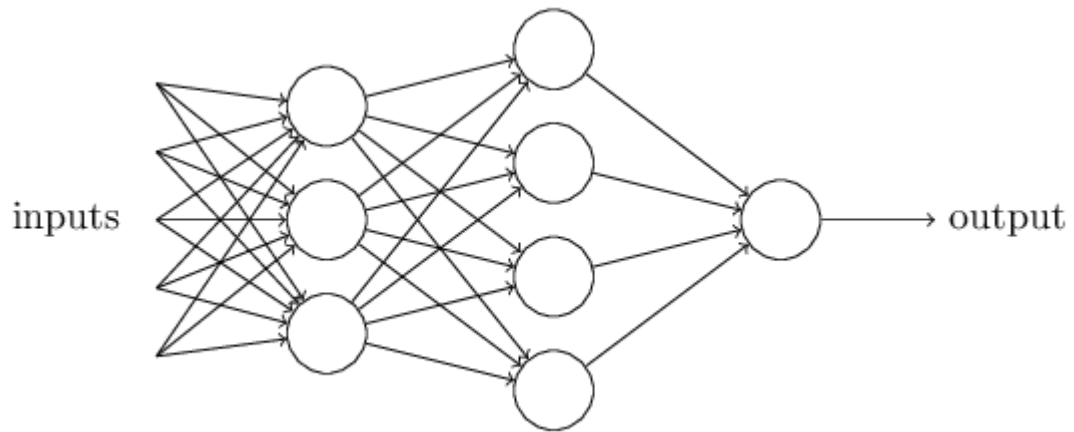


Рисунок 1.6 - Приклад схеми роботи нейронної мережі

Cognitive Computing (Когнітивні обчислення) - ще один варіант штучного інтелекту. Він складається із систем, які виконують завдання або приймають конкретні рішення як помічники або замітники людей, оскільки вони можуть справлятися з двозначністю й невизначеністю та мають високий ступінь автономії в межах своєї сфери знань.

Когнітивні обчислення відносяться до систем, які навчаються в масштабі, міркують цілеспрямовано і природньо взаємодіють з людиною. Замість того, щоб бути явно запрограмованими, вони навчаються і міркують на основі взаємодії з нами і на основі свого досвіду з навколишнім середовищем. Когнітивні системи є імовірнісними. Вони генерують не просто відповіді на числові задачі, але й гіпотези, обґрунтовані аргументи і рекомендації щодо більш складних - і значущих - масивів даних. [8]

Хоча він включає в себе елементи штучного інтелекту, когнітивні обчислення - це ширша ідея, ніж розробка машин, які "думають замість людей". Когнітивні обчислення спрямовані на розширення людського інтелекту та поглиблення знань - допомагаючи кожному краще мислити і приймати обґрунтовані рішення.

Image Segmentation (Сегментація зображення) - це широко розповсюджений метод обробки та аналізу цифрових зображень для поділу зображення на кілька

частин або областей, часто на основі характеристик пікселів у зображенні. Сегментація зображення може передбачати відокремлення переднього плану від фону або кластеризацію областей пікселів на основі подібності кольору чи форми.

Для групування пікселів можна використовувати будь-яку ознаку, яку можна пов'язати з пікселем. Після того, як пікселі згруповані у кластери на основі цих значень вимірювального простору, можна легко знайти пов'язані області за допомогою міток пов'язаних компонентів.

Загальною проблемою кластеризації є розбиття множини векторів на групи зі схожими на групи зі схожими значеннями. В аналізі зображень вектори представляють пікселі або іноді невеликі околиці навколо пікселів.

У традиційній кластеризації є K кластерів C_1, C_2, \dots, C_k із середніми значеннями m_1, m_2, \dots, m_k . Міру похибки за методом найменших квадратів визначають за формулою, яка вимірює, наскільки дані близькі до призначених їм кластерів (рис. 1.7).

$$D = \sum_{k=1}^K \sum_{x_i \in C_k} \|x_i - m_k\|^2.$$

Рисунок 1.7 - Формула для вимірювання міри похибки

Поширеним застосуванням сегментації зображення в медичній візуалізації є виявлення та позначення пікселів у зображенні, які представляють пухлину в мозку або інших органах пацієнта. Алгоритми і методи для сегментації зображень були розроблені протягом багатьох років з використанням предметних знань для ефективного вирішення проблем сегментації в кожній конкретній області застосування. Програми з таким методом обробки даних, корисні для медичної візуалізації, автоматизованого водіння, відеоспостереження та машинного бачення.

Hyperparameter (Гіперпараметр) - це параметр ШІ, який встановлюється перед початком процесу генерації. Ці параметри можна регулювати і безпосередньо впливати на те, наскільки якісним буде результат. Прикладами гіперпараметрів у машинному навчанні є швидкість генерації, кількість отриманих зображень, стиль тощо. Гіперпараметри мають прямий вплив на алгоритми штучного інтелекту. Таким чином, щоб досягти максимальної продуктивності та задовільного результату, важливо чітко розуміти, як їх оптимізувати.

Сам алгоритм навчання часто має гіперпараметри $\lambda \in \Lambda$, а власне алгоритм навчання - це алгоритм, отриманий після вибору λ , який можна позначити A_λ . Обчислення, що виконуються самим A , часто включають в себе "внутрішню оптимізацію" наприклад, оптимізація ваги запиту. Оптимізація гіперпараметрів можна розглядати як задачу "зовнішньої оптимізації".

Її формулюють за спеціальною формулою (рис. 1.8).

$$\lambda^{(*)} = \arg \min_{\lambda \in \Lambda} E_{(x,y) \in (X,Y)} [\mathcal{L}(y, \mathcal{A}_\lambda((X, Y)^{(\text{train})}))].$$

Рисунок 1.8 - Формула оптимізації гіперпараметрів

Text-to-image model (Модель «Текст в зображення») - це модель машинного навчання, яка приймає вхідний опис природною мовою та створює зображення, яке відповідає цьому опису. Моделі перетворення тексту в зображення, як правило, поєднують мовну модель, яка перетворює вхідний текст на приховане зображення, і генеративну модель, яка створює картинку на основі цього прихованого зображення. Найефективніші моделі, як правило, тренувалися на величезній кількості зображень і текстових даних, зібраних з Інтернету.

Основні технології використання штучного інтелекту представлені на рисунку 1.9.

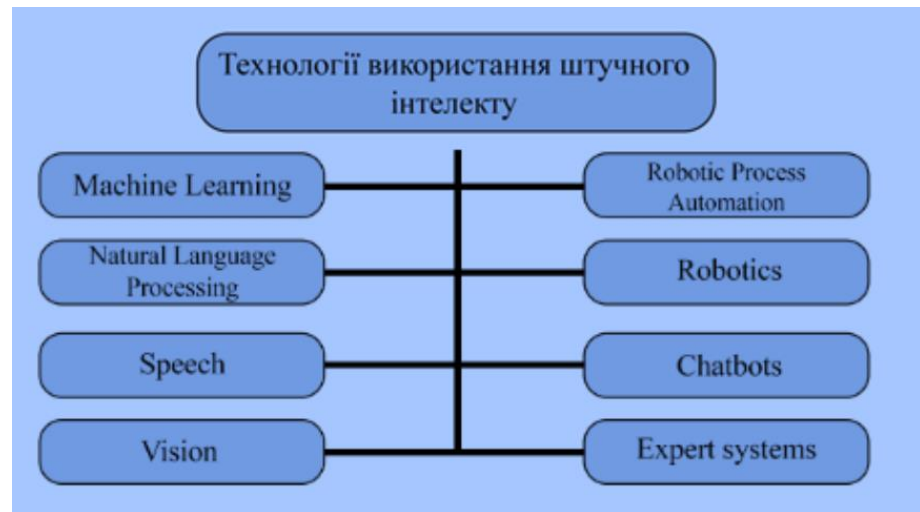


Рисунок 1.9 - Схема технологій використання штучного інтелекту

Технологія генерації зображень за текстом з'явилась відносно нещодавно, але вже повністю змінює роботу багатьох індустрій.

Сучасне дослідження штучного інтелекту розпочалося у 1956 році, коли комп'ютерний вчений Джон Маккарті узгодив зі своїми колегами провести 8-тижневу сесію мозкового штурму в Дартмуті з питання відтворення в машинах різних компонентів людського розуму, таких як мова, зорове сприйняття та мислення. Однією з перших ідей, яка була висунута на конференції, є те, що «Кожний аспект навчання, чи будь-яка інша риса інтелекту в принципі може бути настільки точно описана, щоб можна було створити машину для її імітації». [2]

Період, який наступив після цієї події, став відомий як "Золоті роки ШІ": урядові агенства, почали масово фінансувати дослідження штучного інтелекту. Проте, коли дослідники зрозуміли всі труднощі, стало очевидно, що більшість прогнозів щодо ШІ були далекими від реальності. Обмежена потужність комп'ютерів означала, що вони все ще "занадто слабкі, щоб проявити інтелект". Так званий "комбінаторний вибух" виявив серйозну проблему масштабованості, оскільки для вирішення складніших завдань потрібна була експоненційна кількість вхідних даних.

Дослідники формальної логіки стикалися з великими труднощами, а парадокс Моравека, за яким найважче відтворити в машинах ті здібності людини, які здаються найочевиднішими, наприклад, моторні здібності, все більше підтверджувався. Через всі ці прикросі, в історії ШІ було декілька періодів, які назвали «Зими штучного інтелекту»(рис. 1.10), оскільки через низку невдач у дослідженнях фінансування та зацікавленість суттєво скоротилися.

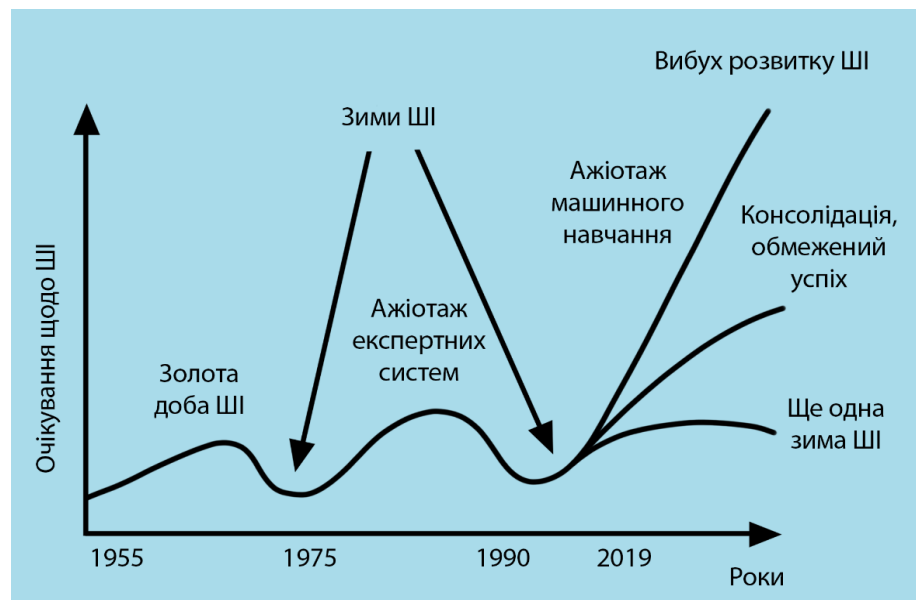


Рисунок 1.10 – Схема зим штучного інтелекту

Найбільш ефективною технологією, що зараз використовується в індустрії штучного інтелекту, зокрема в генерації зображень, є нейронні мережі. [10]

Нейронні мережі (ANN):

Нейронні мережі (ANN) — це модель, натхненна роботою людського мозку. Вони складаються з нейронів, що імітують обчислювальні функції мозку. ANN використовується для завдань класифікації, прогнозування та розпізнавання образів. Великі нейронні мережі, такі як DNN (глибокі нейронні мережі), вирішують складні завдання, включаючи роботу з великими обсягами даних (рис. 1.11).

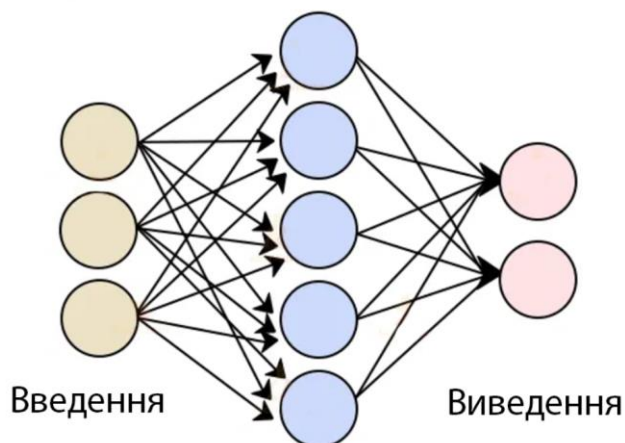


Рисунок 1.11 – Схема роботи моделі ANN

Штучні нейронні мережі (CNN):

CNN створені та оптимізовані для обробки зображень та використовують такий математичний прийом як «згортка». Вони ефективно вирішують завдання з розпізнавання зображень, включаючи застосування у визначенні об'єктів та розпізнаванні облич (рис. 1.12).

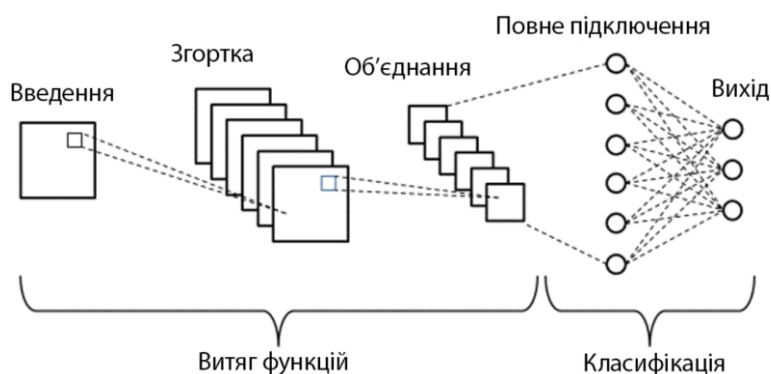


Рисунок 1.12 – Схема роботи моделі CNN

Рекурентні нейронні мережі (RNN):

У рекурентних нейронних мережах вузли, що складаються з нейронів, об'єднуються в цикли. У результаті вихід з одного вузла може вплинути на вхід наступних вузлів. RNN призначені для обробки послідовностей даних, таких як мовлення або рукописні тексти. Вони використовують внутрішню пам'ять, щоб обробляти вхідні дані та зберігати інформацію про попередні етапи обчислень.

RNN ефективно використовується у завданнях, де важливі такі речі, як машинний переклад або генерація тексту (рис. 1.13).

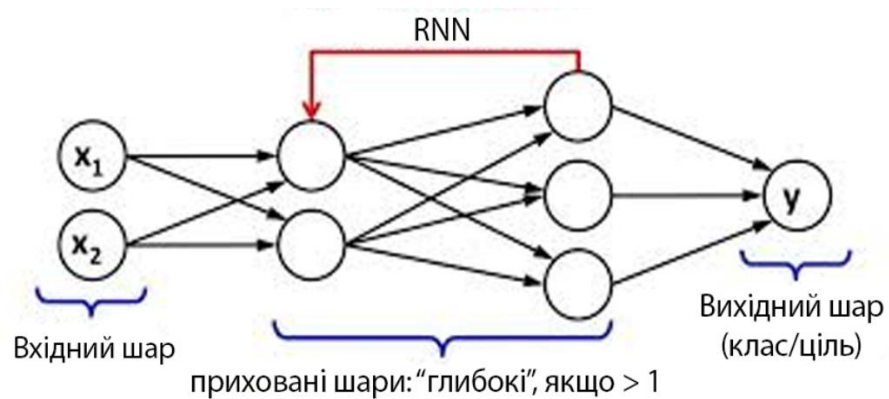


Рисунок 1.13 – Схема роботи моделі RNN

Генеративно-суперечливі мережі (GAN):

Це одна з моделей штучного інтелекту, яка може розпізнавати шаблони у вхідних даних. Генеративно-суперечливі мережі використовують дві моделі — генератор і дискримінатор — для генерації нових даних, що максимально наближаються до вхідних даних. Ці мережі широко використовуються в генерації зображень, відео та обробці голосу(рис. 1.14).

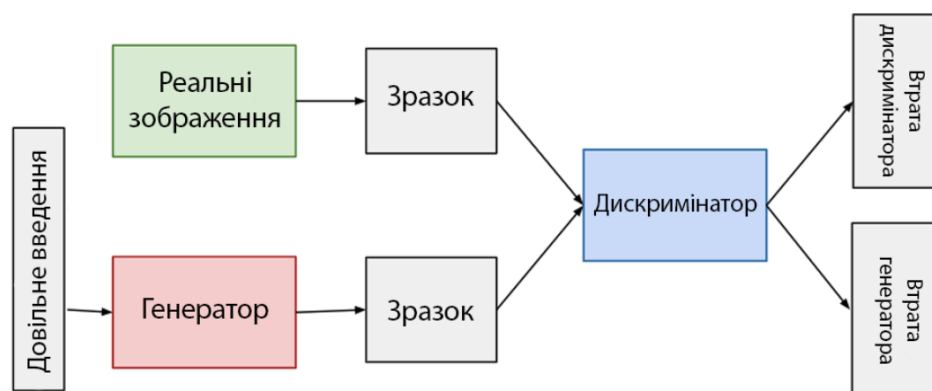


Рисунок 1.14 – Схема роботи моделі GAN

Коріння створення генерації зображень сягає розвитку генеративних моделей, серед яких лідирують генеративні змагальні мережі (GAN) та варіаційні автокодера (VAE). GAN, представлені Яном Гудфеллоу та його колегами у 2014

році, працюють за принципом генеративної моделі, що протистоїть дискримінаційній моделі у змагальному процесі. Цей процес призводить до генерації зображень, які можуть бути реалістичними, імітуючи патерни та особливості навчальних даних.

По суті, генерація зображень передбачає створення візуального контенту системами штучного інтелекту без безпосередньої участі людини. Ці системи можуть створювати нові, часто сюрреалістичні зображення, долаючи розрив між людською уявою та обчислювальними можливостями. Застосування генерації зображень різноманітне - від творчих мистецьких проєктів до доповнення даних у машинному навчанні. Технологія знаходить своє місце у створенні контенту, ігровому дизайні, віртуальних середовищах і навіть у створенні реалістичних облич для аватарів.

Prompts (Підказки/Ідеї) - це письмовий опис, який вказує ШІ Text-to-image (текст в зображення), як йому створити зображення. Штучний інтелект може генерувати зображення, що імітують різні види мистецтва, такі як абстрактні картини маслом, фотореалістичні портрети тощо, залежно від заданих параметрів. Промпти визначають основну тему зображення і також включають в себе всі необхідні деталі, такі як художній матеріал, стиль та композиція (рис. 1.15).

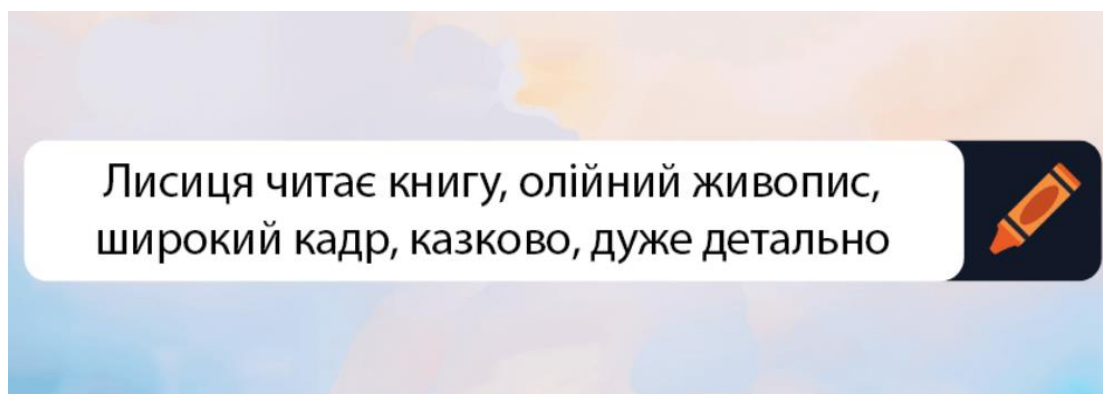


Рисунок 1.15 – Приклад промπτу для ШІ

Застосування промптів в штучному інтелекті для генерації зображень відкриває нові горизонти для творчості користувачів. За допомогою цих підказок, можливості отримання унікальних та вражаючих візуальних елементів стають більш доступними та простими. Продовжуючи розвиватися, системи генерації ШІ на основі промптів впроваджують все більше нових рішень та покращень, роблячи програми ще більш ефективними.

Створення промптів для штучного інтелекту не завжди є систематичним і точним процесом. Потрібно багато практики та велика кількість проб і помилок, щоб знайти правильні слова для створення того самого ідеального бажаного зображення. Головними моментами, яких слід дотримуватися у створенні промптів, є точно визначений стиль, детально описаний основний предмет зображення та середовище навколо нього. Також, для більш деталізованого результату генерації, до підказок можна додати опис джерел світла, бажане кадрування та роздільну здатність, розмиття фону і змінити розмір та співвідношення сторін зображення.

1.2 Проблематика галюцинацій штучного інтелекту в генерації зображень

AI Hallucination (Галюцинації ШІ) - це явище, при якому мовна модель штучного інтелекту, частіше за все генеративний чат-бот або інструмент комп'ютерного бачення, сприймає за дійсність шаблони чи об'єкти, які не існують або непомітні для користувача, створюючи безглузді або зовсім неточні результати.

Як правило, якщо користувач робить запит до ШІ, він бажає отримати результат, який належним чином відповідає промпту (тобто повністю правильна відповідь на запитання). Однак інколи, алгоритми штучного інтелекту видають

результати, які не ґрунтуються на реальних даних, неправильно декодуються або не відповідають жодному шаблону, який можна ідентифікувати. Іншими словами, ШІ видає «галюцинаційну» відповідь.

Більшість наркотиків, які викликають галюцинації — викликають і переживання, під час яких люди бачать, чують, відчують запах, або відчують те, чого не існує. Вони роблять це, впливаючи на мозкові центри “гормонами щастя”. За схожим принципом можна симулювати галюцинації, які називаються "DeepDream" (глибокі сни), надмірно стимулюючи їх.

У результаті нерідко утворюються сноподібні, яскраві, психоделічні зображення(рис. 1.16):



Рисунок 1.16 - Галюцинації нейронних мереж

Цей термін може здатися нелогічним і абсолютно парадоксальним, враховуючи, що галюцинації зазвичай пов’язані з мозком людини чи тварини, а не з машинами. Але з метафоричної точки зору, саме слово «галюцинація» точно описує хибні результати ШІ, особливо у випадку розпізнавання образів і генерації зображень (де результати можуть бути справді дивними та сюрреалістичними на вигляд).

Галюцинації ШІ схожі на те, як люди іноді бачать дивні фігури або чують голоси, яких насправді не існує. У випадку штучного інтелекту, ці неправильні інтерпретації відбуваються через різні фактори, включаючи переобладнання, зміщення/неточність навчальних даних і високу складність запиту.

Всі галюцинації штучного інтелекту є прямим результатом великих мовних моделей (LLM), які дозволяють генеративним інструментам штучного інтелекту

(наприклад ChatGPT) обробляти мову подібно до людини. Незважаючи на те, що LLM розроблені для створення плавного та зв'язного тексту, вони не мають розуміння реальності, яку вони описують. Усе, що вони роблять, це передбачають, яким буде наступне слово, ґрунтуючись на ймовірності, а не на точності.

Щоб зрозуміти, як це відбувається, важливо знати, як працюють LLM. LLM отримують величезну кількість текстових даних, включаючи книги, новинні статті, блоги та публікації в соціальних мережах. Ці дані потім розбиваються на менші одиниці, які називаються токенами, вони можуть бути короткими, як одна літера, або довгими, як слово.

LLM використовують нейронні мережі, щоб з'ясувати, як ці слова та літери працюють разом. Нейронні мережі складаються з блоків обробки, так званих вузлів, які з'єднані один з одним через вагові коефіцієнти. Ці ваги встановлюються шляхом надання моделі певного тексту та її спроби заповнити наступне слово, а потім порівняння результату з тим, що насправді було в тексті. Це відбувається знову і знову, і модель кожного разу коригує свої внутрішні параметри, щоб краще робити ці прогнози, але насправді модель ніколи не дізнається значення самих слів. [9]

Оскільки модель обробляє все більше і більше тексту, вона починає розпізнавати шаблони в мові, такі як правила граматики та асоціації слів, навчаючись розуміти, які слова, ймовірно, слідують за іншими в реченні. З часом модель розвиває форму семантичного розуміння, де вона вчиться пов'язувати слова та фрази з їхніми значеннями. Саме це дозволяє моделі ШІ писати супровідні листи, створювати рецепти, давати поради та виконувати всі інші завдання. Тим не менш, LLM не можуть повністю зрозуміти реальність того, про що вони говорять.

Галюцинації ШІ можуть мати значні наслідки для нашого реального життя. Наприклад, модель штучного інтелекту в сфері охорони здоров'я може неправильно визначити доброякісне ураження шкіри як злоякісне, що призведе до непотрібних медичних втручань. Також, хибні відповіді ШІ можуть сприяти

поширенню дезінформації. Якщо, наприклад, галюцинаційні новинні боти відповідають на запити про надзвичайну ситуацію, що розвивається, неперевіреною інформацією, вони можуть швидко поширити неправду, яка поширить паніку серед людей.

Одним з найбільших джерел галюцинацій в алгоритмах машинного навчання є зміщення вхідних даних. Якщо модель штучного інтелекту навчається на наборі даних, що містить упереджені або нерепрезентативні дані, вона може галюцинувати закономірності або особливості, які відображають ці упередження.

Наразі пропонується новий підхід до проблеми галюцинацій, формулюючи метод виявлення галюцинацій і використання синтетичних галюцинованих текстів для створення набору даних, які можна буде використовувати як основу для майбутніх фільтрів і механізмів, що в кінцевому підсумку можуть стати основою частиною архітектури LLM.

Вихідний матеріал сегментовано за словами, водночас мітка "0" була призначена правильним словам, а мітка "1" - словам-галюцинаціям. Нижче ми бачимо приклад галюцинації, пов'язаної з вхідною інформацією, але доповнену неавтентичними даними (рис. 1.17).

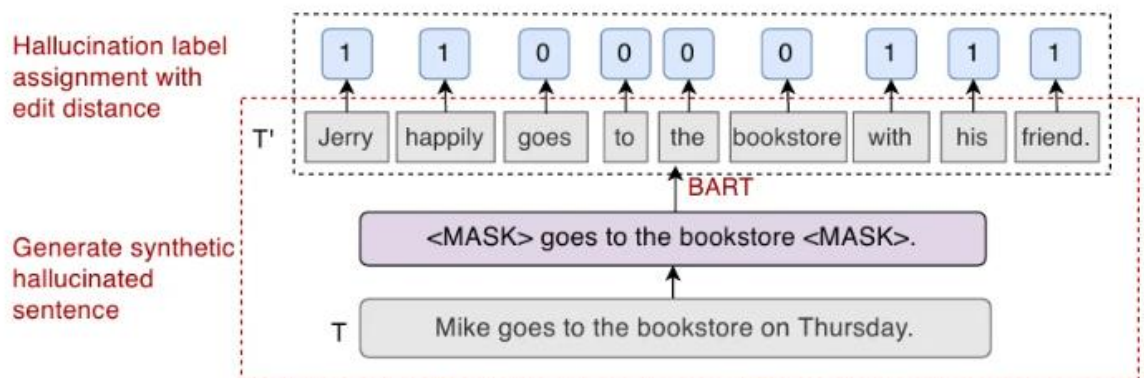


Рисунок 1.17 - Приклад сегментації слів

У системі використовується попередньо навчений шумоподавлюючий автокодувальник, здатний відобразити галюцинаційний рядок назад у вихідний текст, з якого було створено викривлену версію.

Процес зіставлення галюцинації з джерелом, яке неможливе за звичайного використання високорівневих моделей LLM, дає змогу відобразити "відстань редагування" і полегшує алгоритмічний підхід до ідентифікації змісту галюцинацій (рис. 1.18).

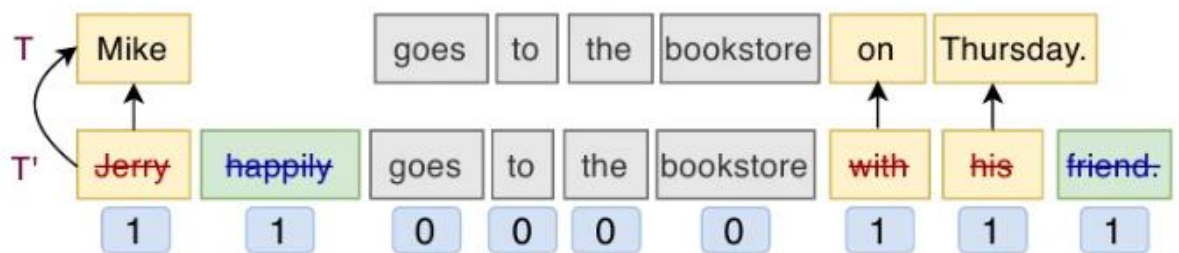


Рисунок 1.18 - Процес зіставлення

Дослідники виявили, що система здатна добре узагальнювати навіть тоді, коли у неї немає доступу до довідкових матеріалів, які були доступні під час навчання, що говорить про те, що концептуальна модель надійна і широко відтворювана.

З розвитком технологій штучного інтелекту, різноманітні ШІ показують дивовижні результати в створенні зображень, роблячи їх неймовірно якісними та майже невідмінними від людської творчості. Однак цей процес супроводжується не тільки успіхом і час від часу стикається з рядом складнощів та невдач. Більш точні формулювання запиту та підказки можуть зробити результати кращими, але є речі, з якими поки що ШІ не має змоги впоратися.

Експерти з штучного інтелекту борються з цими проблемами, але це вимагає довгих годин уточнень та підказок для ШІ, або ретушування створених зображень за допомогою різних фоторедакторів. Тому, на даний момент, якщо знати, на що звертати увагу при дослідженні згенерованого зображення, то є великий шанс, що ви таки зможете зрозуміти, чи було воно створене машиною.

Найбільш відомою проблемою ШІ зі створенням зображень є неправильна генерація рук людини, в результаті чого виходить «каша рук» або «куля з пальців». Хоч розробники штучного інтелекту досягли достатньо непоганих результатів з удосконалення генерації зображень, та все ще існує багато речей, які можна покращити. Один з найкращих загальнодоступних генераторів зображень DALL-E від OpenAI створює зображення людей, які на перший погляд виглядають дуже реалістично, проте поглянувши на їхні руки, стають очевидними деякі проблеми штучного інтелекту (рис. 1.19).



Рисунок 1.19 – Приклади генерації рук від ШІ

Також, однією з найпоширеніших проблем штучного інтелекту є проблеми з текстом на зображеннях. Це може здаватися дивним, адже вважається, що генерація тексту є дуже легкою задачею, але насправді, для реалістичного

зображення потрібно враховувати стиль, кут і перспективу розміщення, а також освітлення.

У цьому прикладі відносно новий генератор зображень Leonardo AI спробував створити дизайн придорожного кафе у вінтажному стилі та з назвою «Jack Rabbit Slim's». Як ми бачимо, вінтажний стиль дійсно помітний, картинка в цілому виглядає добре, але з текстом маємо явні проблеми (рис. 1.20).



Рисунок 1.20 – Проблеми з текстом у створенні зображень

Окрім попередніх проблем, також існують певні труднощі з генерацією очей та зубів (рис. 1.21).



Рисунок 1.21 – Приклад генерації людей штучним інтелектом

Очі можуть бути однією з найважливіших деталей у зображенні портретів, але багатьом ШІ не вдається коректно відобразити очі людини на малюнках. Штучний інтелект Bing Image Creator досить непогано впорався з генерацією сімейного фото, проте можна помітити великі проблеми з очима людей (рис. 1.22).



Рисунок 1.22 – Проблеми з зображенням очей

Коли люди посміхаються, це зазвичай робить картинку більш приємною, але штучний інтелект може перетворити таке звичне явище як посмішка в справжнє жахіття (рис. 1.23).



Рисунок 1.23 – ШІ генерує зуби людини

1.3 Дослідження етичних проблем, пов'язаних з використанням штучного інтелекту

За всією своєю потужністю та перспективами впливу на різні аспекти нашого життя, штучний інтелект вносить і низку етичних проблем, які вимагають глибокого вивчення та розгляду. В цьому розділі, присвяченому етичним проблемам ШІ, ми вдосконалимо наше розуміння та усвідомимо етичні виклики, які супроводжують цей швидкозмінний технологічний прорив в науковому світі.

Розглядаючи ШІ як інструмент, який може об'єднувати і розвивати людство, але при цьому породжує серйозні етичні питання, ми віддамо належну увагу конфліктам інтересів, які виникають між різними сторонами в контексті застосування штучного інтелекту. Також ми детально розглянемо важливість забезпечення прозорості, конфіденційності та правильного визначення відповідальності у випадках використання ШІ.

Першою і найочевиднішою етичною проблемою, є розповсюдження шкідливого контенту. Системи штучного інтелекту можуть створювати листи і подібні речі автоматично на основі текстових підказок людей. Електронний лист, створений штучним інтелектом і надісланий від імені компанії, наприклад, може випадково містити образливу лексику або давати шкідливі вказівки працівникам. Генеративний штучний інтелект слід використовувати як допоміжну технологію, а не для заміни людей чи процесів, щоб забезпечити відповідність вмісту етичним очікуванням компанії та підтримувати її цінності бренду. [6]

Другою проблемою є порушення авторського права та юридична незахищенність. Популярні генеративні інструменти штучного інтелекту навчаються на величезних базах даних зображень і тексту з багатьох джерел, включаючи Інтернет. Коли ці інструменти створюють зображення або генерують рядки коду, джерело даних може бути невідомим, що може бути проблематичним, наприклад для банку, який обробляє фінансові операції або фармацевтичній компанії, яка покладається на формулу складної молекули в ліках. Репутаційні та фінансові ризики також можуть бути величезними, якщо продукт однієї компанії базується на інтелектуальній власності іншої. Компанії повинні перевіряти результати генерації штучним інтелектом, доки юридичні прецеденти з проблемами інтелектуальної власності та авторських прав існують.

Наступною етичною проблемою, яку ми розглянемо, є порушення конфіденційності даних. Генеративні моделі штучної мови ШІ навчаються на наборах даних, які іноді включають особисту інформацію про людей. Іноді, ці дані можна отримати за допомогою простої текстової підказки. Порівняно з традиційними пошуковими системами, людині, чия конфіденційність постраждала завдяки ШІ, може бути важче знайти інформацію та вимагати її видалення. Компанії, які створюють або налаштовують такі моделі штучного інтелекту, повинні переконатися, що ідентифікаційна інформація не вбудована в мовні моделі ШІ та що її можна легко видалити з цих моделей відповідно до законів про конфіденційність.

Генеративний ШІ демократизує можливості штучного інтелекту і робить їх доступнішими для всіх. Таке поєднання демократизації та доступності, потенційно може призвести до того, що медичний дослідник ненавмисно розкриє конфіденційну інформацію про пацієнта, або споживчий бренд мимоволі розкриє свою стратегію просування продукту третій стороні. Наслідки ненавмисних інцидентів подібних до цих, можуть безповоротно порушити довіру пацієнтів або клієнтів і мати правові й матеріальні наслідки. Компаніям рекомендується запровадити чіткі вказівки, управління та ефективну комунікацію згори вниз, наголошуючи на спільній відповідальності за захист конфіденційної інформації, захищеності даних та інтелектуальної власності.

Далі ми розглянемо таку проблему, як посилення існуючих стереотипів та упереджень. Ця проблема стає важливою у зв'язку з тим, як інтелектуальні системи ШІ вивчають та аналізують величезні обсяги даних, включаючи відомості, що містять стереотипи та упередження.

Однією з основних причин цієї проблеми є те, що інтелектуальні системи навчаються на основі існуючих даних, які можуть відображати соціокультурні стереотипи, що існують у суспільстві. Коли алгоритми навчаються на даних, де існують упередження щодо різних груп людей за такими ознаками, як раса, стать, вік та інші, інтелектуальна система може виробити та поширити аналогічні судження.

Наприклад, у системах відеоспостереження, які використовують розпізнавання облич, може виникнути проблема стереотипізації певних груп людей, що може призвести до неправильних висновків або визначень. Також, у сфері відбору кандидатів на роботу за допомогою інтелектуальних систем, виникнення стереотипів може призвести до несправедливого відбору та призначення осіб.

Рішення цієї етичної проблеми потребує системного підходу. Зокрема, важливо активно впроваджувати методи контролю, перевірки та аудиту алгоритмів

на предмет стереотипів. Крім того, застосування диверсифікованих та репрезентативних даних для навчання інтелектуальних систем може допомогти зменшити вплив стереотипів.

Штучний інтелект повинен однаково допомагати всьому суспільству, не розповсюджуючи та не посилюючи упередження. Ефективне вирішення цієї проблеми визначатиме, наскільки ШІ внесе позитивний вплив на різнобічність, рівність та справедливість в сучасному світі.

Останньою етичною проблемою, яку ми розглянемо, буде заміна людей штучним інтелектом. Це є однією з ключових тем, яка породжує численні етичні, соціальні та економічні питання в епоху швидкої автоматизації та розвитку інтелектуальних систем. Ця дискусія заслуговує уваги, оскільки вплив ШІ на ринок праці та життя людей стає все більш суттєвим.

Однією з основних проблем є втрата робочих місць через автоматизацію та роботизацію в різних секторах економіки. ШІ, обладнаний високою продуктивністю та ефективністю, може замінити людей у багатьох рутинних роботах, що може призвести до безробіття та втрати основних засобів до існування для багатьох робочих категорій людства.

Інша проблема полягає в нерівномірному впливі на різні галузі і суспільні групи. Якщо автоматизація впливає на традиційні сфери праці, то це може призвести до дисбалансу та соціально-економічної нерівності. Також, можливість виникнення "цифрового розколу" може далі розширити розрив між тими, хто може користуватися перевагами ШІ, і тими, кого це не стосується.

Етичні питання також виникають у контексті взаємодії ШІ з людьми. Якщо інтелектуальні системи штучного інтелекту можуть приймати різноманітні важливі рішення, чи можемо ми довіряти їхнім рішенням у сферах, таких як медицина, право та соціальна справедливість? Як забезпечити, щоб рішення ШІ були справедливими, прозорими та відповідали етичним стандартам?

Заміна людей ШІ може мати і позитивні сторони, такі як автоматизація небезпечних або монотонних завдань, але необхідно уважно вивчати та регулювати цей процес, забезпечуючи збалансоване співіснування технологій та людського потенціалу для досягнення гармонійного розвитку.

Отже, приймаючи до уваги все вищесказане, можна зробити висновки, що стрімко рухаючись у світле майбутнє, де штучний інтелект сприяє загальному благу, ми несемо відповідальність за формування цього майбутнього таким чином, щоб воно було етично збалансованим і справедливим для всіх.

2 ДОСЛІДЖЕННЯ ТЕХНОЛОГІЧНИХ РЕСУРСІВ ГЕНЕРАЦІЇ ЗОБРАЖЕНЬ

2.1 Сучасне програмне забезпечення для генерації зображень

Одним з перших сучасних ШІ для створення зображень, стало програмне забезпечення DeepDream від Google, що виникло в глибокій згортковій мережі під кодовою назвою «Початок» і було випущене в 2015 році.

Нейронні мережі, такі як DeepDream, мають біологічні аналогії, що дають уявлення про роботу мозку і формування свідомості. Нейронні мережі навчаються на вхідних векторах і змінюються під впливом внутрішніх варіацій під час процесу навчання. Вхідні та внутрішні зміни відображають обробку екзогенних та ендогенних сигналів відповідно у зоровій корі. Оскільки внутрішні варіації змінюються в глибоких нейронних мережах, вихідне зображення відображає ці зміни. Ця специфічна маніпуляція демонструє, як внутрішні механізми мозку аналогічні внутрішнім шарам нейронних мереж.

Алгоритм DeepDream ініціює процес, пересилаючи певне зображення через мережу, а потім починає вимірювати градієнт зображення щодо певного шару активації. На наступному кроці зображення налаштовується, щоб покращити ці активації та підсилити шаблони, які призводять до картинки, схожої на сон. Весь цей процес також відомий як інцепціонізм.

Нейронна мережа отримує зображення як вхідні дані та надає результат аналізу. Це досягається шляхом розпізнавання ознак в ієрархічному порядку. Перші нейрони, які отримують прямий доступ до зображення, як правило, чутливі до дуже простих характеристик, таких як певні кольори, форми та положення. Потім наступний шар нейронів отримує вхідні дані від попереднього та створює функції вищого рівня. Частина об'єктів, як правило, розпізнаються на цьому рівні: дверні ручки, ніжки столу тощо. Постійно додаючи шар поверх попереднього, нейронна мережа може розробляти все більш складні функції, доки зрештою не розпізнає різницю між, наприклад, обличчям kota і людини (рис. 2.1).

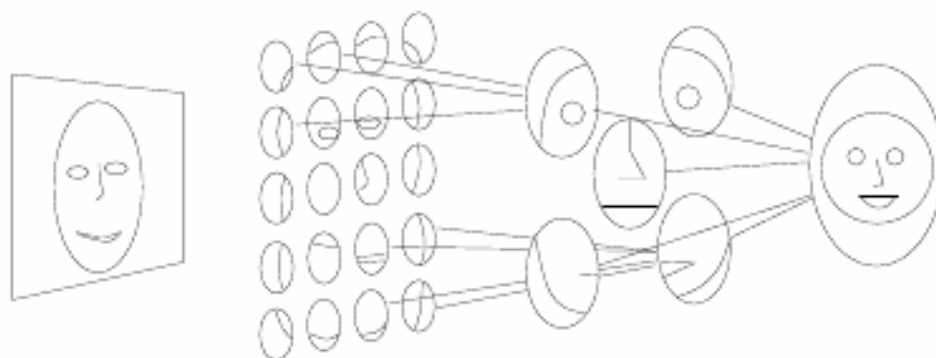


Рисунок 2.1 - Процес виділення ієрархічних ознак

У додатку 1 наведено приклад коду Python, який використовується для обчислень DeepDream.

Наступним ШІ, який ми розглянемо, стане DALL-E від розробників ChatGPT. Цю модель анонсували 5 січня 2021 року, а вже 6 квітня 2022 року було анонсовано наступну версію - DALL-E 2. Це моделі ШІ з перетворення тексту в зображення, розроблені OpenAI з використанням методології глибокого навчання для створення цифрових зображень із описів природною мовою, які називаються “промптами”(підказками). У вересні 2023 року OpenAI анонсувала свою останню версію ШІ, DALL-E 3, здатну розуміти «значно більше нюансів і деталей», ніж попередні ітерації.

Модель DALL-E є мультимодальною реалізацією GPT-3 із 12 мільярдами параметрів, які «замінюють текст на пікселі», навчені на парах текст-зображення з Інтернету. Вхідними даними для моделі є послідовність символізованих підписів, за якими слідують маркеровані фрагменти зображень. Підпис до зображення англійською мовою, токенизоване кодуванням пари байтів (розмір словника 16384) і може містити до 256 токенів. Кожне зображення — це зображення RGB розміром 256x256, розділене на 32x32 ділянки 4x4 кожна. Потім кожен патч перетворюється дискретним варіаційним автокодером на маркер (розмір словника 8192).

DALL-E було розроблено та анонсовано разом із CLIP (попередня підготовка контрастної мови й зображення). CLIP — це окрема модель, заснована на нульовому навчанні, яка була навчена на 400 мільйонах пар зображень із текстовими підписами, взятими з Інтернету. Ця модель використовується для фільтрації більшого початкового списку зображень, згенерованих DALL-E, щоб вибрати найбільш відповідні результати.

Contrastive Language-Image Pre-training — це методика навчання пари моделей. Одна модель приймає фрагмент тексту та виводить один вектор. Інша приймає зображення та виводить один вектор.

Щоб навчити таку пару моделей, можна почати з підготовки великого набору даних пар зображення-підпис, а потім вибірки партій розміру N . Нехай результати текстової та графічної моделей будуть відповідно $U_1, \dots, U_n, W_1, \dots, W_n$. Втрати складатимуть (рис. 2.2):

$$-\sum_i \ln \frac{e^{v_i \cdot w_i}}{\sum_j e^{v_i \cdot w_j}} - \sum_j \ln \frac{e^{v_j \cdot w_j}}{\sum_i e^{v_i \cdot w_j}}$$

Рисунок 2.2 - Загальна сума втрат крос-ентропії в кожному стовпці та кожному рядку матриці

Випущені моделі були навчені на наборі даних "WebImageText", що містить 400 мільйонів пар підписів до зображень. Загальна кількість слів подібна до WebText, який містить близько 40 ГБ тексту.

DALL-E може створювати зображення в різних стилях, включаючи фотореалістичні зображення, картинки та емодзі. Він може "маніпулювати та переставляти" об'єкти на своїх зображеннях і може правильно розміщувати елементи дизайну в нових композиціях без чітких інструкцій.

Враховуючи наявне зображення, DALL-E 2 може створювати «варіації» зображення, як окремі картинки на основі оригіналу, а також редагувати зображення, щоб змінити або розширити його. Функції «Зафарбувати» і «Розфарбувати» у DALL-E 2 використовують контекст із зображення, щоб заповнити відсутні ділянки за допомогою середовища, сумісного з оригіналом, дотримуючись заданої підказки.

Третім дослідженим ШІ стане Midjourney, один з найпопулярніших наразі штучних інтелектів для створення зображень. Ця платформа була заснована в Сан-Франциско і вперше увійшла у відкриту бета-версію 12 липня 2022 року.

Midjourney наразі доступний лише через бота на офіційному сервері Discord, шляхом прямого обміну повідомленнями або запрошенням бота на сторонній сервер. Основними, для роботи з цим ШІ, є три команди:

“/imagine” - основна команда, після якої вводять потрібний промпт і отримують набір з 4 зображень.

“/blend” - дозволяє користувачу змішати два зображення.

“/shorten” - підказує, як зробити промпт коротшим та більш результативним.

Оскільки Midjourney має закритий код, ми не можемо точно знати, як саме працює цей ШІ, але можемо описати базову технологію.

Midjourney покладається на дві відносно нові технології машинного навчання, а саме на великі мовні моделі та дифузійні моделі. З першою моделлю ми знайомі з використання генеративних чат-ботів ШІ, наприклад ChatGPT. Велика мовна модель спочатку допомагає Midjourney зрозуміти значення слів, які ми вводимо у підказках. Потім це перетворюється на так званий вектор, який можна уявити, як числову версію підказки. Нарешті, цей вектор допомагає керувати іншим складним процесом, відомим як дифузія.

У моделі дифузії комп'ютер поступово додає випадковий шум до свого навчального набору зображень. З часом він навчиться відновлювати вихідне

зображення, змінюючи шум. Ідея полягає в тому, що після достатнього навчання така модель може навчитися створювати абсолютно нові зображення.

Отже, розглянемо як це виглядає з точки зору генератора зображень.

Коли ми вводимо текстову підказку, вона починається з поля візуального шуму, як еквівалент телевізійної статики. Далі навчена модель штучного інтелекту використовує зворотню дифузію для поступового віднімання шуму. Згодом це дасть картину, яка зображує об'єкти та ідеї в реальному світі, та відповідає заданому промпту.

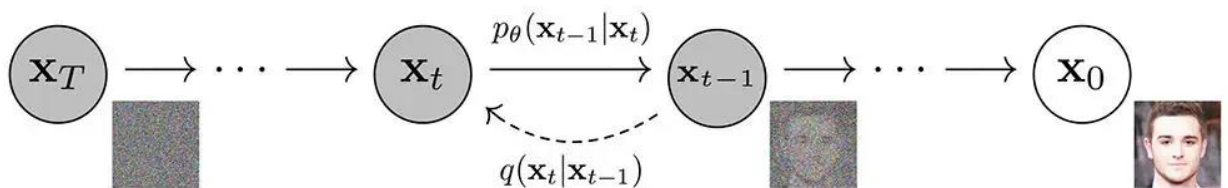


Рисунок 2.2 - Процес зворотної дифузії

2.2 Порівняння функціональних можливостей програмного забезпечення для генерації зображень

Штучний інтелект в генерації зображень став дивовижним каталізатором для розширення горизонтів та вдосконалення творчих можливостей. Різноманітні моделі та алгоритми призначені для різних завдань, всі вони мають свої унікальні характеристики, які визначають те, як вони справляються з різними творчими завдання і порівнювати різні ШІ важливо для розуміння їхніх сильних та слабких сторін.

У цьому дослідженні та порівняльному аналізі ми детально розглянемо кілька ключових ШІ для генерації зображень: DreamStudio, Midjourney, DALL-E 2.

Кожен з цих ШІ має власний набір переваг та недоліків, що впливають на їхню здатність створювати реалістичні, естетичні та інноваційні зображення.

Наша мета - визначити, які з цих штучних інтелектів вигідно використовувати для конкретних завдань, розібравши та порівнявши їхню ефективність, можливості та обмеження. Давайте розглянемо ці моделі детальніше, звертаючи увагу на їхню унікальність, результативність та вплив на світ генерації зображень.

Першим ШІ, який ми розглянемо, буде достатньо популярна програма DreamStudio, яка використовує останню версію моделі генерації зображень Stable Diffusion (модель стабільної дифузії) (рис. 2.3).

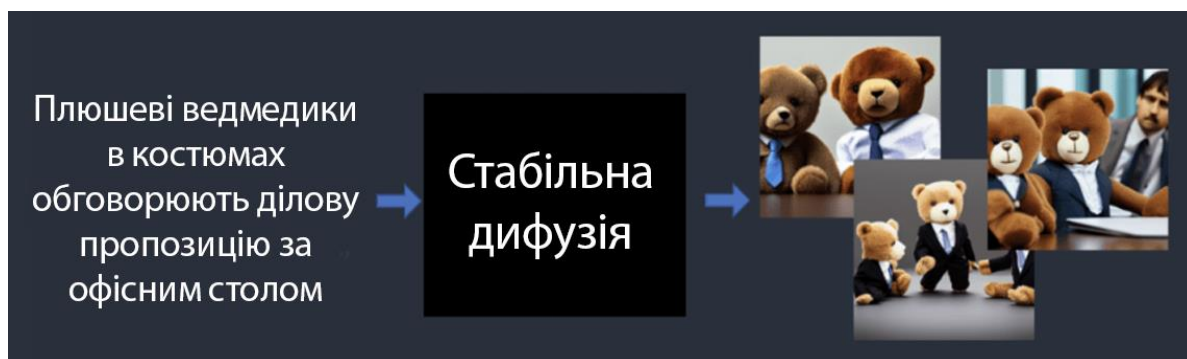


Рис. 2.3 – Стабільна дифузія перетворює текст на зображення

Цей штучний інтелект має відкритий код та постійно вдосконалюється й оновлюється. Він здатний створювати будь-які зображення, від найпростіших ілюстрацій до реалістичних фотографій і справжніх витворів мистецтва.

DreamStudio проста у використанні та пропонує багато функцій для налаштування. Наприклад, окрім введення основного запиту до зображення, ви можете вибрати стиль генерації (наприклад фотографічний або 3D-модель), встановити широту і висоту картинки, визначити кількість згенерованих зображень тощо (рис. 2.4).

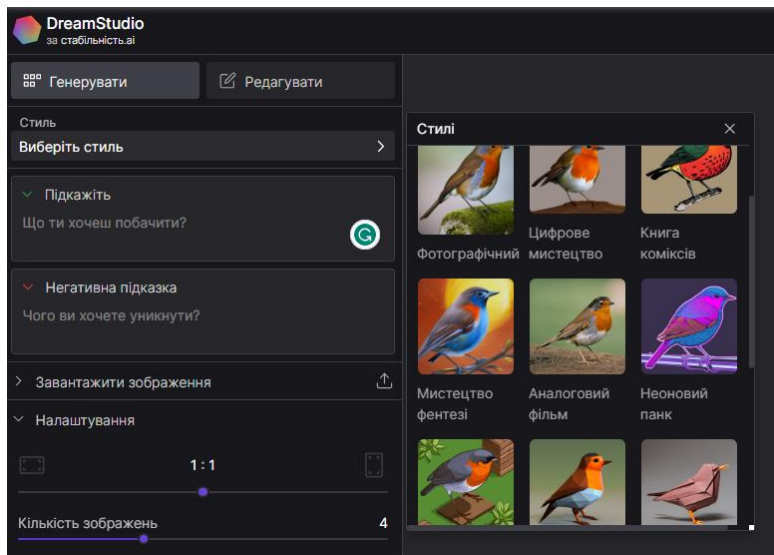


Рис. 2.4 – Меню налаштувань DreamStudio

Тепер, давайте спробуємо згенерувати декілька картинок та проаналізуємо якість зображень, які створює цей ШІ. DreamStudio безперечно вміє створювати реалістичні зображення, а деякі результати дійсно вражають.

Ввівши запит “A realistic photo of a red cat in a red beret sitting near the Eiffel Tower, with people in the background” ми отримали дійсно якісний результат, який справді важко відрізнити від реальної фотографії, зробленої людиною (рис. 2.5).



Рисунок 2.5 – Результат першого запиту в DreamStudio

Вдруге, запит на генерацію звучав так «Two women stand dressed in futuristic black suits, with a night city in the background» і результат вийшов не таким успішним. В цілому, картинка непогана, але ми можемо помітити явні проблеми з генерацією людей, а саме, дивні спотворені обличчя. Можна зробити висновки, що для хорошого зображення треба буде зробити декілька спроб, але загалом, отримати бажаний результат більш ніж можливо. (рис. 2.6)



Рисунок 2.6 – Результат другого запиту в DreamStudio

Також, спробувавши інший режим генерації, а саме “Ілюстрацію”, ми отримали дуже хороший результат, який дійсно виглядає як намальована людиною картина. Запит для штучного інтелекту був таким - «A view of Paris at night in the style of Van Gogh's painting "Starry Night"» (рис. 2.7).



Рисунок 2.7 - Результат третього запиту в DreamStudio

Щодо ціноутворення DreamStudio, то після реєстрації на сайті користувачі можуть безкоштовно отримати 25 балів, що дозволяє декілька разів згенерувати зображення. Після того, як витратили ці бали, можна придбати додаткові, ціна починається від \$10 за 1000 балів.

Другим ШІ, який ми дослідимо, стане Midjourney. Це ще один дуже популярний генератор зображень, який більше підходить для продвинутих користувачів, які мають більший досвід у створенні ілюстрацій за допомогою штучного інтелекту. Платформа Midjourney пропонує різні моделі штучного інтелекту, масштабування зображень, поєднання стилів та додаткові параметри для генерації.

Midjourney став першим штучним інтелектом, який взяв участь у мистецькому конкурсі і переміг. В 2022 році в США проводився конкурс, на якому перше місце посіла картина з назвою “Théâtre D’opéra Spatial”, яку художник Д. Аллен створив за допомогою ШІ (рис. 2.8). [12]



Рис. 2.8 – Картина Midjourney “Théâtre D’opéra Spatial”

Суттєвим недоліком цього ШІ є те, що його використання доступне тільки в соціальній мережі Discord, тобто окремого сайту у нього немає. Користувачу потрібно витратити деякий час на вивчення цієї соціальної мережі, а потім ще й додатковий час на всі налаштування та можливості Midjourney. Всі побажання до генерації зображення, наприклад стиль та розміри потрібно писати безпосередньо в повідомленні з основним запитом.

Ще одним мінусом Midjourney є те, що всі згенеровані зображення також з'являються в спільноті Discord Midjourney, що викликає справедливі занепокоєння щодо конфіденційності використання цього штучного інтелекту. Хоча, з іншого боку, це допомагає бачити прогрес інших користувачів, навчатися у них та знаходити натхнення в чужих роботах. Незважаючи на недоліки цього ШІ, його результати генерації зображень варті уваги. Midjourney може створювати дуже реалістичні зображення їжі, пейзажів та людей і видає кращі результати, якщо запит сформульований якомога детальніше (рис. 2.9).

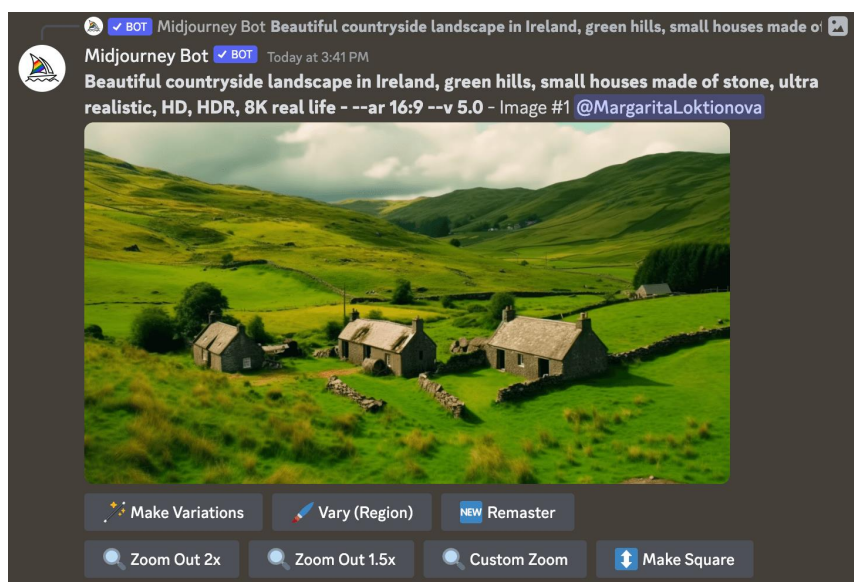


Рис. 2.9 – Меню генерації Midjourney в Discord

Першим запитом знову був “A realistic photo of a red cat in a red beret sitting near the Eiffel Tower, with people in the background”. Результат виглядає трохи менш

реалістичним за зображення від DreamStudio, але все ж картинка вийшла деталізованою та реалістичною (рис. 2.10).



Рисунок 2.10 – Результат першого запиту до Midjourney

Другий запит для Midjourney звучав так - “Very realistic dish, rice with salmon, photography, high detail, HD, HDR, real life, 16:9 format”. Результатом стало дуже гарне зображення, яке неможливо відрізнити від справжньої фотографії (рис. 2.11).



Рисунок 2.11 – Результат генерації Midjourney

Незважаючи на те, що Midjourney використовувати складніше за альтернативні програми, цей ШІ пропонує неймовірні можливості для створення унікальних і справді реалістичних зображень.

Третім запитом для цього штучного інтелекту був “Bird's eye view photo of Manhattan at night, the city is very bright, lots of light, skyscrapers, HD, HDR, real life, 16:9 format” (рис. 2.12).



Рисунок 2.12 – Результат генерації третього запиту

Також, Midjourney успішно може впоратися з будь-якими іншими стилями мистецтва, наприклад ілюстрацією тощо.

Щодо ціноутворення на цей ШІ, то базова підписка коштує \$10 на місяць і дає можливість створити близько 200 зображень. Окрім базової, є більш дорогі версії підписки, з якими можна згенерувати ще більше картинок.

Третім ШІ, який ми дослідили, став DALL-E 2 від розробників всім відомого ChatGPT, який є одним з найвідоміших генераторів зображень. Він дозволяє створювати зображення з будь-якими візуальними ефектами, а також має декілька

цікавих функцій, таких як редактор зображень, гумку тощо. За допомогою гумки можна замінювати елементи на згенерованих ШІ картинках.

З мінусів цього штучного інтелекту – помітно гірша деталізація та генерація зображень людей . Хоча розробники і обіцяють, що майбутня DALL-E 3 позбавиться цих недоліків і матиме розширені можливості для генерації, та поки що маємо, що маємо.

Перший запит до DALL-E 2 є ідентичним до запиту DreamStudio, а саме - «A realistic photo of a red cat in a red beret sitting near the Eiffel Tower, with people in the background». Результат, хоч і непоганий, але дійсно гірший за зображення DreamStudio, менш реалістичний і має недоліки (рис. 2.13).



Рисунок 2.13 – Результат першого запиту до DALL-E 2

Загалом, DALL-E 2 більше підходить для створення цифрового мистецтва та ілюстрацій, ніж для реалістичних фотографій.

Другим запитом до цього штучного інтелекту став «A view of Paris at night in the style of Van Gogh's painting "Starry Night"». Результатом стало, хоч і не деталізоване, але по своєму цікаве й незвичне зображення (рис. 2.14).



Рисунок 2.14 – Результат другого запиту до DALL-E 2

Щодо ціни, то підписка для цього ШІ коштує \$15 і дає змогу створити близько 460 зображень.

2.3 Методи виявлення штучно згенерованих зображень

Завдяки своїй здатності імітувати людський інтелект і створювати надзвичайно реалістичний контент, ШІ став потужним інструментом, який може формувати інформацію та маніпулювати неї. Оскільки контент, створений штучним інтелектом, стає все більш поширеним, виникає важливе питання: як ми можемо відрізнити реальне від штучно створеного? Щоб ідентифікувати зображення та переконатися в їх автентичності, можна використовувати різні методи.

Виявлення контенту, створеного за допомогою ШІ, є надзвичайно важливим у сучасному цифровому середовищі. З розвитком технології штучного інтелекту зловмисникам стало дедалі легше створювати фейкові новини, спам або навіть маніпулювати громадською думкою. Неправдивий контент, створений штучним інтелектом, може вводити в оману користувачів, завдаючи значної шкоди окремим

особам і суспільству в цілому. Застосовуючи ефективні методи виявлення, ми можемо захиститися від поширення неправдивої або оманливої інформації.

Такі методи, як аналіз стилю, виявлення шаблонів або застосування алгоритмів машинного навчання, можуть допомогти ідентифікувати створений ШІ контент і пом'якшити його негативний вплив. Виявлення контенту, створеного штучним інтелектом, дає нам кращі можливості для підтримки автентичності та цілісності інформації, що просувається в Інтернеті.

Методи виявлення штучно згенерованих зображень:

Аналіз текстур та патернів - використовує алгоритми для визначення характерних текстур та патернів на зображеннях, здатний виявляти аномальні області, які можуть вказувати на синтетичне походження.

Відстеження цифрових слідів - аналізує метадані та цифрові сліди, такі як дата та час зйомки, параметри камери тощо, перевіряє їхню узгодженість з очікуваними характеристиками для реальних фотографій.

Аналіз моделей глибокого навчання - вивчає структуру та параметри моделей глибокого навчання, що використовуються для генерації зображень, виявляє особливості, які можуть свідчити про їхнє штучне походження.

Взаємодія з користувачем - залучає користувачів для позначення або перевірки автентичності зображень.

Використання Metadata та Exif-даних - аналізує додаткові дані, такі як Exif-дані та інші метадані, що зберігаються в файлі зображення, виявляє відмінності у метаданих, які можуть свідчити про синтетичне походження, пошук невидимих або зашифрованих слів та позначень.

Метрики реалізму - використовує числові метрики для визначення реалістичності зображень, порівнює різні характеристики, такі як освітлення, тіні, колір та композиція.

Системи визначення глибокого фейку - використовує спеціалізовані системи для виявлення глибоких фейків та штучно створених зображень, аналізує ключові ознаки, характерні для глибоких фейків.

Методи виявлення штучно згенерованих зображень стають все більш важливими у світлі зростаючої популярності та доступності глибокого навчання. Вирішення цієї проблеми визначається не лише технічними можливостями, але й соціальним та етичним контекстом. Забезпечення безпеки та достовірності візуальної інформації вимагає співпраці між дослідниками, розробниками технологій та суспільством в цілому. Розвиток та вдосконалення методів виявлення є кроком у напрямку забезпечення цифрової екосистеми, яка враховує та захищає візуальну правдивість та довіру.

2.4 Висновки

В останні роки ШІ для генерації зображень зробили суттєвий прорив та стали катализатором для вдосконалення розуміння творчості, дозволяючи за лічені хвилини створити фантастичні роботи, які дійсно вражають своєю реалістичністю. Ці інноваційні технології відкривають нам нові горизонти, дозволяючи людям експериментувати та розкривати свій творчий потенціал.

Кожна з досліджених програм зі штучним інтелектом має свої переваги та недоліки, роблячи досвід роботи з ними по справжньому унікальним та неповторним. Різноманітність інструментів ШІ дозволяє кожному знайти саме той варіант, який врахує всі вимоги та зможе задовольнити унікальні потреби кожної людини.

Зараз складно зробити точні прогнози, щодо перспектив розвитку і вдосконалення технологій ШІ, але перед нами точно відкривається шлях до абсолютно нових досягнень в творчості. З розвитком алгоритмів та удосконаленням моделей штучного інтелекту, ми можемо розраховувати на ще більшу точність та результативність в творчості ШІ.

3 ПРАКТИЧНА РЕАЛІЗАЦІЯ

3.1 Дослідження та генерація зображень в Midjourney

Першим ШІ, який ми обрали для проведення експериментів, стало програмне забезпечення Midjourney.

Штучний інтелект Midjourney знаходиться на передовій технологічних інновацій, використовуючи потужність алгоритмів машинного навчання і нейронних мереж, щоб переосмислити сферу обробки та генерації зображень. Ця передова система штучного інтелекту може похвалитися безліччю переваг, використовуючи свою складну архітектуру для забезпечення безмежної продуктивності в різних сферах застосування.

Однією з ключових переваг ШІ Midjourney є глибокі згорткові нейронні мережі, які дозволяють моделі розпізнавати складні патерни та особливості зображень. Глибина цих мереж забезпечує високий рівень абстракції, що дозволяє Midjourney розуміти складні візуальні ієрархії з надзвичайною точністю.

Крім того, ШІ Midjourney оснащений розширеними можливостями трансферного навчання, що дозволяє йому використовувати попередньо навчені моделі на великих наборах даних. Це значно прискорює процес навчання для конкретних завдань, пов'язаних із зображеннями, таких як розпізнавання облич, аналіз медичних зображень і промисловий контроль якості. Трансферне навчання дозволяє моделі адаптувати та узагальнювати свої знання для нових областей, тим самим зменшуючи потребу у великих маркованих наборах даних і прискорюючи розгортання рішень ШІ.

Отже, перейдімо до експериментів з промптами та досліджень того, як їхнє формулювання відображається на результатах генерації. Починаючи від найпростіших текстових промптів, до більш складних і деталізованих, буде показана очевидна необхідність прописання чітких вимог для генерації.

Для першого зображення, ми використали дуже короткий, нечіткий, та не деталізований промпт. Він звучав так:

“Make a beautiful representation of a mountain landscape”

Результат генерації представлено на рисунку 3.1

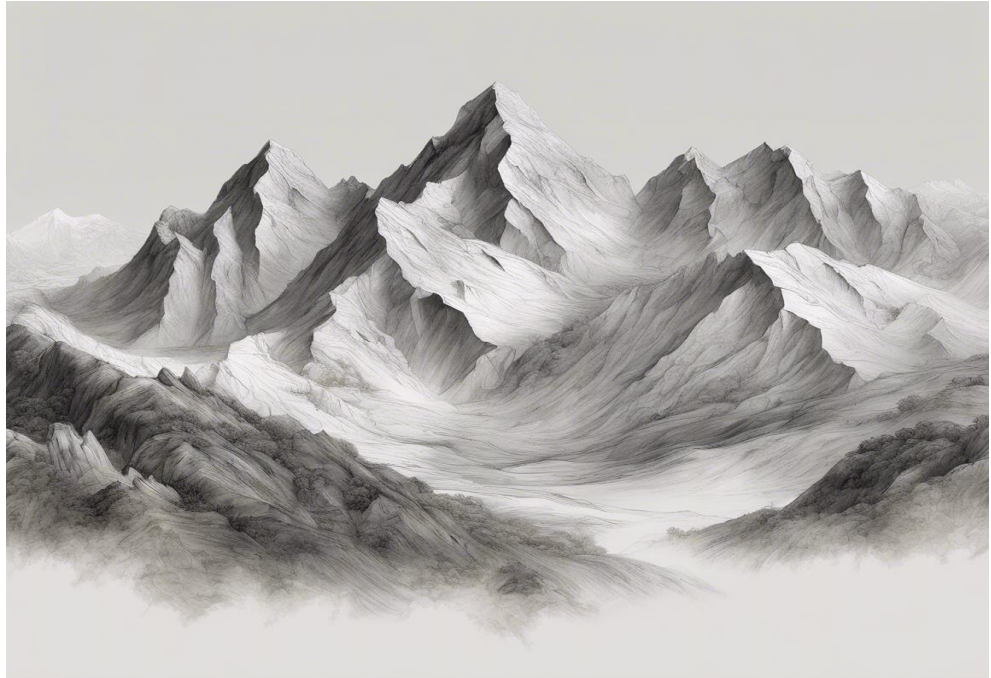


Рисунок 3.1 - Результат генерації першого промпту в Midjourney

Отримавши занадто просте зображення, робимо висновок, що треба покращити наш промпт та додати більше елементів. Покращений промпт звучить так:

“Generate a detailed image of a peaceful mountain scene at sunrise. Picture a landscape with snowy peaks, a meadow, and a stream. Include a small group of sheep grazing in the meadow. Show rocky cliffs and waterfalls”

Результат генерації представлено на рисунку 3.2



Рисунок 3.2 - Результат генерації другого промпту в Midjourney

Згенероване зображення помітно краще та має багато цікавих деталей, проте все ще виглядає достатньо просто. Отримавши такі результати, ми приходимо до висновку, що потрібно написати промпт, у якому чітко будуть розписані абсолютно всі елементи майбутнього зображення. Третій варіант промпту звучить так:

“Generate a stunning and highly detailed image capturing the breathtaking beauty of a mountain landscape during the golden hours of dawn. Envision a panoramic scene where majestic snow-capped peaks rise dramatically against a pastel-hued sky, transitioning from hues of soft lavender to warm tinges of peach and amber as the sun begins its ascent. In the foreground, envision a lush alpine meadow adorned with an exquisite array of wildflowers, their delicate petals catching the first light of the day. The meandering path of a crystal-clear mountain stream adds a touch of serenity, reflecting the vibrant colors of the sunrise. Nestled within this idyllic landscape, imagine a small flock of sheep peacefully grazing, their woolly forms scattered across the meadow. Picture the sunlight gently illuminating their pristine white fleece, casting subtle shadows as they contentedly nibble on the verdant grass. The sheep exude a sense of tranquility, seamlessly blending into the harmonious tapestry of the mountain scene. Zoom in to

capture the intricate details of the sheep, highlighting the individual strands of their wool, the subtle textures of their fur, and the curiosity in their expressive eyes. Each blade of grass should be meticulously rendered, swaying in the mountain breeze, conveying a sense of realism that transports the viewer into this pristine natural setting. As the landscape ascends towards the towering peaks, introduce craggy rock formations, each weathered by the passage of time and the elements. Snow patches cling to the higher altitudes, glistening in the morning sunlight, while waterfalls cascade down the rugged cliffs, creating a symphony of natural sounds that reverberate through the mountainous terrain. Include diverse vegetation, from hardy evergreen trees clinging to the slopes to delicate alpine flowers that carpet the ground. Illuminate the scene with dynamic lighting, emphasizing the interplay of light and shadow across the varied terrain, accentuating the contours of the mountains and enhancing the overall depth of the composition. Consider incorporating a subtle haze that adds an ethereal quality to the distant peaks, creating an atmospheric perspective that enhances the grandeur of the mountain landscape. Super resolution, ProPhoto RGB, Backlighting, Cinema Lighting, insanely detailed and intricate, hyper maximalist. High detail, HD, HDR, real life, 5:4 format”

Результат генерації представлений на рисунку 3.3



Рисунок 3.3 - Результат генерації третього промпту в Midjourney

Завдяки цьому промπτу, ми отримали необхідне зображення, на генерацію пішло не більше хвилини та видно кожен деталь, описану в текстовій підказці. Отже, генеративний штучний інтелект Midjourney є вражаючим інструментом у сфері генерації зображень, демонструючи можливості перетворення текстових підказок у деталізовані твори мистецтва.

Як було помічено, коли промпти деталізовані та містять багато елементів, містять чіткі інструкції щодо елементів сцени, умов освітлення та складних об'єктів, ШІ Midjourney реагує на них з підвищеним рівнем точності та достовірності. Здатність моделі вловлювати та враховувати найдрібніші деталі, від текстури овечої вовни до гри світла на гірській місцевості, особливо помітна, коли підказки деталізовані.

І навпаки, менш детальні підказки можуть дати більш узагальнені результати, оскільки модель покладається на свої навчальні дані та загальні знання для інтерпретації та генерації зображень. Важливість точності підказок стає очевидною для досягнення бажаного рівня реалістичності та конкретності згенерованих зображень.

3.2 Дослідження та генерація зображень в DALL-E

Другим ШІ, який було обрано для проведення експериментів, стало програмне забезпечення DALL-E. Цей штучний інтелект відомий передусім тим, що був створений в OpenAI, одній із найамбітніших у світі лабораторій штучного інтелекту.

DALL-E може створювати зображення різноманітних довільних описів з різних точок зору, та рідко допускає помилок. Також було виявлено, що DALL-E може поєднувати концепції, а його здібностей до візуального мислення достатньо для розгадування матриць Рейвена.

Однією з ключових переваг DALL-E є його здатність генерувати зображення широкого спектру концепцій, від повсякденних об'єктів до сюрреалістичних і абстрактних сценаріїв. Модель демонструє розуміння просторових зв'язків і може генерувати візуально цілісні зображення, які відповідають текстовим підказкам. Ця універсальність робить DALL-E цінним інструментом для творчих починань, дозволяючи користувачам візуалізувати концепції в новий і візуально вражаючий спосіб.

На жаль, платформа DALL-E надає дуже обмежену кількість безкоштовних спроб для генерації зображень, а саме 5 спроб на один обліковий запис. Окрім цього, цей штучний інтелект не має додаткових налаштувань, тобто все, на що користувач може вплинути під час генерації, це текстовий промпт.

Отже, почнімо наше дослідження генерації зображень за допомогою DALL-E. Як і з Midjourney, починаючи експерименти від найпростішого промпту, до складного й багатослівного, ми прослідкуємо прив'язаність результату ШІ до деталізації текстових підказок.

Перший промпт для генерації буде звучати так:

“Create an image of a seascape at sunset”

Зображення, що вийшло в результаті генерування, показано на рисунку 3.4.



Рисунок 3.4 - Результат генерації першого промпту в DALL-E

Оскільки результат має не дуже багато деталей та виглядає не реалістично, спробуємо краще розписати текстову підказку. Другим промптом буде:

“Generate an image portraying a serene seascape at sunset. Picture a vast coastal landscape with sandy shores and rugged rocks, the horizon blending seamlessly with the tranquil sea. The sky transforms with warm hues—deep oranges, purples, and pinks—scattered with wispy clouds catching the last rays of the setting sun”

Результат другого промпту показано на рисунку 3.5



Рисунок 3.5 - Результат генерації другого промпту в DALL-E

Ми отримали якісну хорошу картинку, та все ж, їй не вистачає більшості елементів та реалізму. Отже, зробимо промпт, який ще точніше опише наше бачення майбутнього зображення, та всіх його аспектів. Третій промпт звучатиме так:

“Create a mesmerizing image capturing the serene beauty of a seascape at sunset. Picture a vast coastal landscape, where the merging elements of earth, sky, and water paint a breathtaking canvas. As the sun begins its descent, envision a sprawling coastline adorned with a harmonious blend of sandy shores and rugged rocks. The horizon stretches infinitely, forming a seamless connection between the earthly realm and the vastness of the tranquil sea. The sky above transforms into a spectacular symphony of colors, embracing the warm embrace of the setting sun. Vivid shades of deep orange, purple, and pink dance across the firmament, creating a celestial masterpiece. Delicate, wispy clouds catch the final rays of daylight, their forms intricately woven into the canvas of the sky, adding depth and dimension to the scene. As the sun sinks lower, its radiant glow

intensifies, casting a golden sheen upon everything it touches. The entire seascape bathes in this ethereal light, invoking a sense of tranquility and wonder. Shadows lengthen, and the landscape becomes a play of light and dark, showcasing the contours of the shoreline and the gentle undulations of the sea. The water reflects the celestial spectacle above, mirroring the kaleidoscope of hues that paint the sky. The sea is calm, its rhythmic waves whispering secrets to the shore. A serene atmosphere envelopes the entire scene, inviting contemplation and reflection. In the foreground, imagine the subtle details of the coastal environment—a scattering of seashells on the sand, a lone piece of driftwood weathered by the tides, and perhaps the distant silhouette of a sailboat on the horizon. These elements add a touch of realism and narrative to the composition, inviting viewers to immerse themselves in the tranquil beauty of this seascape at sunset. In the synthesis of this image, aim for a seamless integration of colors, textures, and lighting that authentically captures the magic of a coastal sunset. Let the image evoke a sense of awe and appreciation for the natural wonders that unfold when day gracefully gives way to night along the shores of the sea. High detail, HD, HDR, real life”

Отримане зображення, завдяки третьому промпту, представлено на рисунку 3.6



Рисунок 3.6 - Результат генерації третього промпту в DALL-E

Як ми можемо бачити, завдяки точному опису бажаного зображення, ШІ зміг створити дуже реалістичну картинку та додати до неї майже все описане в промпті.

Можна зробити висновки, що DALL-E є свідченням швидкого прогресу в галузі штучного інтелекту, демонструючи конвергенцію розуміння мови і синтезу зображень. Його здатність генерувати різноманітні, контекстуально релевантні зображення на основі текстового введення відкриває захоплюючі можливості в різних галузях. З розвитком досліджень у цій галузі синтез тексту і зображень може призвести до створення ще більш досконаlih моделей, розширюючи межі можливого в контенті, створеному штучним інтелектом.

3.3 Дослідження та генерація зображень в DreamStudio

Третім ШІ для дослідження обрали DreamStudio на базі технології Stable Diffusion, назва якої походить безпосередньо від методу роботи.

Перевага цього штучного інтелекту над попередніми моделями полягає саме у використанні методу стабільної дифузії. Цей метод гарантує, що згенеровані зображення будуть високої якості і точно відповідатимуть наданому промπτу. За допомогою DreamStudio можна очікувати широкий спектр результатів - від реалістичних портретів і пейзажів до абстрактних композицій.

Також, на відміну від Midjourney та Dall-E, DreamStudio доступний з відкритим вихідним кодом. Це означає, що будь-хто може взяти його основу і безкоштовно створювати додатки, орієнтовані на конкретні завдання з перетворення тексту в зображення. Саме доступність коду DreamStudio, робить справжній прорив у сфері штучного інтелекту, оскільки стане основою для незліченних додатків, вебсайтів і сервісів, які переосмислять те, як ми створюємо та взаємодіємо з мистецтвом.

DreamStudio наразі вважається найкращою генеративною моделлю штучного інтелекту. Цей штучний інтелект не тільки кращий, швидший та простіший у використанні за більшість ШІ, але й дешевший, ніж його конкуренти. У DreamStudio пропонуються безкоштовні кредити, які можна використовувати для створення зображень. Ці кредити корисні для нових користувачів, яким потрібно дослідити інструмент і вивчити його функції, не хвилюючись про миттєву оплату.

Також, DreamStudio надає доступ до бібліотеки, де можна додавати різні стилі до зображень, які користувачі бажають створити. Ця додаткова функція дозволяє генерувати певний результат, не описуючи його в промπτі. Окрім всього, DreamStudio має одну важливу відмінну рису, якої не вистачає багатьом інструментам ШІ подібного характеру - негативні промпти. Завдяки ним, користувач може задати параметри, яких хоче уникнути на згенерованому зображенні.

Отже, проведемо дослідження можливостей штучного інтелекту DreamStudio. В цьому випадку, ми так само рухатимемося від найпростіших текстових промπτів, до більш складних і деталізованих, щоб показати переваги

прописання чітких вимог для генерації. Також, для всіх спроб генерації в DreamStudio, ми будемо використовувати негативний промпт “blur, unrealistic, fantasy”, щоб обмежити появу таких характеристик на зображеннях, а стилем генерації обрали “Enhance”.

Першим промптом для генерації стало просте речення:

“A medieval castle in a forest, towering trees, a path”

Отримавши 4 варіанти генерації зображення, ми обрали найвдаліший результат, який представлений на рисунку 3.7

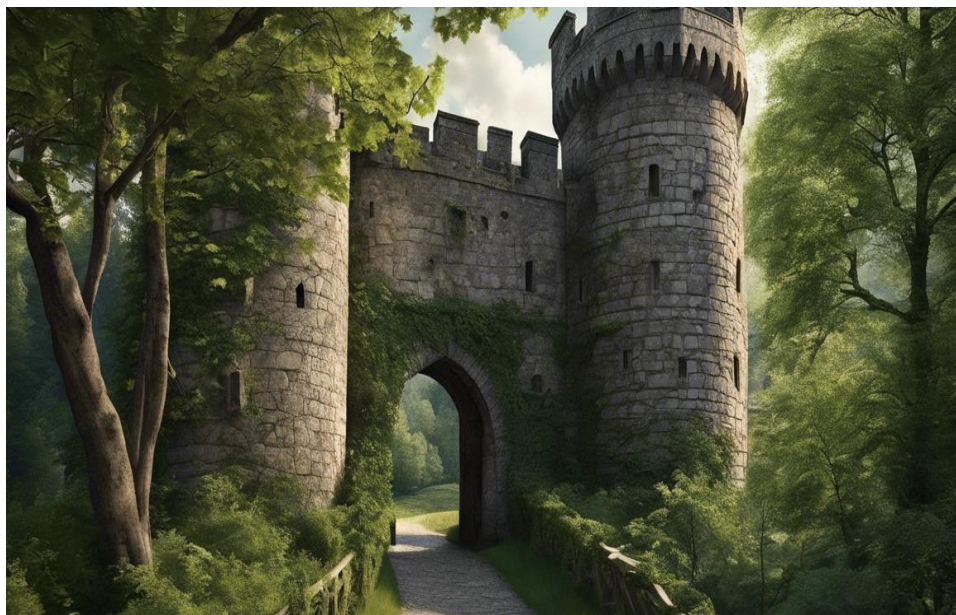


Рисунок 3.7 - Результат генерації першого промпту в DreamStudio

Як ми бачимо, зображення вийшло не реалістичним, простим, та невідповідним заданій підказці, тому, створимо складніший промпт, який звучатиме так:

“An image portraying a medieval castle seamlessly integrated into a lush forest. Envision towering trees forming a natural canopy, a meandering path leading to imposing castle gates. Picture moss-covered stones, ferns, and a crystal-clear moat with a drawbridge. The castle, weathered and ivy-clad, stands as a timeless sentinel”

В результаті, ми отримали зображення, представлене на рисунку 3.8



Рисунок 3.8 - Результат генерації другого промпту в DreamStudio

Зображення стало більш деталізованим, додалися нові елементи, проте всеодно недостатньо. Тому, було вирішено дописати промпт, щоб він максимально точно описував всі необхідні нам подробиці, в результаті чого отримали таку текстову підказку:

“Picture a sprawling image featuring a medieval castle nestled amidst the heart of an enchanting forest. Let your imagination weave a vivid tapestry where the dense foliage embraces the ancient stronghold, creating an atmosphere that is both mysterious and awe-inspiring. The forest, a lush symphony of greens, is punctuated by towering trees whose gnarled branches reach towards the sky, forming a natural canopy that partially conceals the castle within its leafy embrace. As the image unfolds, envision a meandering path leading through the thick undergrowth, guiding the viewer towards the castle's imposing gates. The pathway, dappled with sunlight filtering through the dense foliage, adds an ethereal quality to the journey, beckoning explorers to unravel the secrets hidden within

the heart of the forest. Moss-covered stones and ferns carpet the forest floor, contributing to the ancient and timeless ambiance. The castle itself stands as a formidable structure, its weathered stone walls adorned with ivy, attesting to centuries of existence. Turrets rise majestically against the backdrop of towering trees, their crenellations offering a panoramic view of the surrounding woodland. Gargoyles perch on the corners, silently watching over the secrets held within the castle's walls. Approaching the castle, imagine a drawbridge spanning a crystal-clear moat fed by a babbling brook that winds its way through the forest. A sense of tranquility permeates the scene, with the soothing sounds of nature complementing the medieval architecture. The castle's windows, adorned with leaded glass, glint in the filtered sunlight, hinting at the grandeur within. Delve into the details of the castle courtyard, where a well-worn cobblestone path leads to an arched entrance. Vibrant wildflowers peek through the crevices, adding a touch of color to the otherwise ancient and imposing structure. A lone flag flutters atop a tower, bearing the emblem of a forgotten noble house, further enhancing the narrative of a castle steeped in history. In crafting this masterpiece, infuse the image with a sense of timelessness and mystery. Let the interplay of light and shadow highlight the intricate details of both the natural surroundings and the medieval architecture. The resulting image should transport viewers to a world where the past and present converge, inviting them to explore the secrets held within the medieval castle nestled in the heart of the enchanting forest. High detail, HD, HDR, real life, 5:4 format”

Згенероване за допомогою цього промпту зображення, представлено на рисунку 3.9



Рисунок 3.9 - Результат генерації третього промπτу в DreamStudio

З результату генерації, ми можемо дійти до висновків, що штучний інтелект DreamStudio добре впорався з промπτом та продемонстрував здатність втілювати в життя складні деталі. DreamStudio сплїв воєдино елементи природи та архітектури, створивши зображення, яке відповідає заданому промπτу, хоч і не на всі 100 відсотків. На жаль, десятками спроб генерації ми так і не змогли досягти зображення, яке б повністю відповідало поставленій задачі та зберегло б в собі всі описані деталі та подробиці.

Незважаючи на недоліки, у підсумку, ми бачимо, що штучний інтелект DreamStudio виявився вражаючим інноваційним інструментом для втілення творчих задумів, який встановлює нові стандарти для поєднання технологій та мистецтва в сучасному світі.

ВИСНОВКИ

У цій дипломній роботі ми розглянули технологію штучного інтелекту, дізналися його історію, дослідили широкі можливості ШІ та заглибилися у практичне застосування згенерованих зображень у різних сферах.

В першому розділі представлено детальний огляд історії створення штучного інтелекту. Від його зародження до сучасності, дослідили те, чого штучний інтелект досягнув. Також, у розділі описано складний взаємозв'язок між розвитком ШІ та постійними викликами, які супроводжують його стрімкий прогрес.

Другий розділ ретельно заглиблює в незліченні можливості штучного інтелекту. Ретельно зібравши та проаналізувавши дані про різноманітність моделей штучного інтелекту, ми змогли визначити сильні та слабкі сторони різних ШІ. Цей порівняльний аналіз висвітлив нюанси ШІ, які допомагають з кращим розумінням деталей, пов'язаних з використанням потенціалу цього інструменту.

Третій розділ присвячений практичній реалізації, де теоретичні знання зустрічаються з реальним застосуванням. За допомогою серії експериментів, продемонстрували необхідність чітких та детально сформульованих промптів. Розділ практичного застосування став доказом безмежних можливостей ШІ для генерації зображень.

ПЕРЕЛІК ПОСИЛАНЬ

1. Contributors to Wikimedia projects. History of artificial intelligence - Wikipedia. URL: https://en.wikipedia.org/wiki/History_of_artificial_intelligence
2. The Artificial Intelligence Revolution: Part 1 - Wait But Why. URL: <https://waitbutwhy.com/2015/01/artificial-intelligence-revolution-1.html>
3. A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. (2006). AI Magazine Volume 27, Number 4
4. History of natural language processing https://en.wikipedia.org/wiki/History_of_natural_language_processing
5. Russell, Norvig (2003), Artificial Intelligence: A Modern Approach (2nd ed.)
6. McCorduck (2004), Machines Who Think (2nd ed.), Natick, MA: A. K. Peters, Ltd., ISBN 978-1- 56881-205-2, OCLC 52197627
7. - Elgammal (2017), CAN: Creative Adversarial Networks, Generating “Art” by Learning About Styles and Deviating from Style Norms, <https://arxiv.org/pdf/1706.07068.pdf>
8. Dettmers (2015), The Brain vs Deep Learning Part I: Computational Complexity — Or Why the Singularity Is Nowhere Near
9. Gatys, Ecker, Bethge (2015), A Neural Algorithm of Artistic Style, <https://arxiv.org/pdf/1508.06576.pdf>
10. Progress in artificial intelligence https://en.wikipedia.org/wiki/Progress_in_artificial_intelligence
11. Grace (2018), When Will AI Exceed Human Performance? Evidence from AI Experts, <https://arxiv.org/pdf/1705.08807.pdf>
12. Kevin Roose (2022) An A.I.-Generated Picture Won an Art Prize. Artists Aren't Happy, <https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html>
13. Sukis (2018), The Relationship between Art and AI, <https://medium.com/design-ibm/the-role-ofart-in-ai-31033ad7c54e>

14. An MIT Press book Ian Goodfellow and Yoshua Bengio and Aaron Courville <https://www.deeplearningbook.org/>
15. Robotics, Vision and Control <https://link.springer.com/book/10.1007/978-3-642-20144-8>
16. Speech and Language Processing (3rd ed. draft) <https://web.stanford.edu/~jurafsky/slp3/>
17. Computer Vision: Algorithms and Applications, 2nd ed. <https://szeliski.org/Book/>
18. Hyperparameter Tuning for Machine and Deep Learning with R <https://library.oapen.org/bitstream/id/240dc394-c1a4-414e-8465-d9dc8824c7e5/978-981-19-5170-1.pdf>

ДОДАТОК 1

```
import tensorflow as tf
import numpy as np
import PIL.Image
import urllib.request

# Завантажте попередньо навчену модель InceptionV3 із TensorFlow Hub
base_model = tf.keras.applications.InceptionV3(include_top= False , weights= 'i
base_model.trainable = False

# Визначити список шарів, націлених для візуалізації функцій
layer_names = [ 'mixed3' , 'mixed5' , 'mixed7' ]

# Визначити модель, яка виводить значення активації для цільових шарів
dream_model = tf.keras.Model(inputs =base_model. input , outputs=[base_model.ge

def calc_loss ( img, model ):
    # Обчислити активації цільових шарів
    img_batch = tf.expand_dims(img, axis= 0 )
    layer_activations = model(img_batch)

    losses = []
    for act in layer_activations:
        # Для кожної активації максимізуйте середнє значення активації
        loss = tf.math.reduce_mean(act)
        losses.append(loss)

    return tf.reduce_sum(losses)

def deepdream ( img, model, steps= 100 , step_size= 0.01 ):
    для кроку в діапазоні (кроків):
        # Розрахувати градієнти зображення відносно втрат
        за допомогою tf.GradientTape() як стрічки:
        tape.watch(img )
        .
        -
        -
```

```

- -
-
_ * step_size
img = tf.clip_by_value(img, - 1 , 1 )

return img

def run_deepdream ( image_path, steps= 100 , step_size= 0.01 ):
# Завантажити вхідне зображення
img = tf.keras.preprocessing.image.load_img(image_path)
img = tf.keras.preprocessing.image.img_to_array(img)
img = tf.keras.applications.inception_v3.preprocess_input(img)

# Перетворення на тензор і запуск DeepDream
img = tf.convert_to_tensor(img)
img = deepdream(img, dream_model, steps, step_size)

# Постобробка вихідного зображення
img = tf.keras.preprocessing.image.array_to_img(img)
img .save( 'deepdream_output.jpg' )

# Приклад використання
, якщо __name__ == "__main__" :
image_url = "URL_TO_YOUR_IMAGE"
urllib.request.urlretrieve(image_url, "input_image.jpg" )
run_deepdream( "input_image.jpg" , steps= 100 , step_size= 0,01 )

```