

# Проект GenAI Sketch-to-Image

## GenAI

### Опис проекту

Цей проект спрямований на використання генеративного штучного інтелекту (GenAI) для перетворення ручних ескізів у реалістичні зображення. Користувачі можуть малювати ескізи у веб-інтерфейсі, надсилати їх для обробки та отримувати детальний опис зображення і реалістичну версію свого ескізу. Проект демонструє використання сучасних інструментів ШІ для створення креативного контенту.

<https://sketch-image-lpuu.onrender.com/>

### Функціонал

- **Завантаження ескізів:** Веб-інтерфейс для створення та завантаження ескізів.
- **Опис зображення:** Автоматична генерація детальних описів зображень за допомогою OpenAI API.
- **Перетворення ескізів у зображення:** Трансформація ескізів у реалістичні зображення за допомогою WorqHat API.
- **Обробка зображень:** Ефективна обробка та збереження завантажених і згенерованих зображень.
- **Інтерактивний інтерфейс:** Зручний інтерфейс з миттєвим зворотним зв'язком.

### Вступ

#### Ціль проекту

Метою цього проекту є розробка системи, яка на основі користувацьких ескізів генерує реалістичні зображення та детальні описи цих зображень. Поєднання можливостей сучасних генеративних моделей для обробки зображень і тексту дає змогу легко переходити від концептуального начерку до повноцінної візуалізації.

#### Мотивація для вибору теми

1. **Креативний процес:** Багато ідей та концептів народжуються у вигляді швидких начерків, однак перехід від ескізу до реалістичного зображення зазвичай вимагає тривалого часу та додаткових навичок у графічних редакторах. Автоматизована система допомагає спростити й прискорити цей процес.
2. **Доступність технології:** Завдяки сучасним інструментам генеративного ШІ (наприклад, моделей трансформації зображень та мовних моделей) створення подібних продуктів стало значно простішим. Це відкриває широкі можливості для дизайнерів, ілюстраторів і навіть користувачів без спеціальних навичок.

3. **Потенціал для розширення:** Технологію можна використати в різних сферах – від прототипування UI/UX та створення концепт-арту до освітніх застосунків, де учні можуть одразу бачити результат своїх ідей. Ця гнучкість робить проєкт актуальним та корисним.

## Процес роботи

### Збір і підготовка даних

У даному проєкті не використовувався попередньо зібраний датасет для навчання чи валідації. Усі ескізи надходять безпосередньо від користувача в режимі реального часу. Відповідно, не проводилася додаткова анотація чи фільтрація зображень. Натомість, був підготовлений набір підказок (промптів), які система використовує для опису ескізів та генерування готових зображень.

### Навчання або адаптація моделі

Оскільки в проєкті застосовуються готові API для опису та трансформації зображень (зокрема WorqHat API), процес навчання або fine-tuning власної моделі не проводився. Система повністю покладається на можливості наявних генеративних моделей, доступних через API.

### Інтеграція рішення в продукт або прототип

Усі компоненти були відразу об'єднані в єдиний веб-застосунок. Архітектура передбачає повний цикл роботи з ескізами: від завантаження або малювання користувачем у фронтенд-інтерфейсі, через передачу даних до бекенду для опису та генерації зображень, і до повернення згенерованого результату користувачеві. Окрім основного функціоналу, був реалізований окремий модуль для збору й обчислення метрик (наприклад, CLIP Similarity), що дозволяє оцінити якість та точність згенерованих зображень.

## Виклики та їх вирішення

### Вартість генерації

- **Проблема:** Використання зовнішніх сервісів для аналізу та генерації зображень безпосередньо впливає на бюджет проєкту. Кожен запит до сервісу “Analyze Images” коштує \$0.006, а перетворення “Sketch to Image” – \$0.02 за зображення.
- **Підхід до подолання:**
  - *Оптимізація використання API:* Використання ефективної логіки на рівні бекенду, щоб не надсилати зайві запити та мінімізувати кількість потрібних звернень.

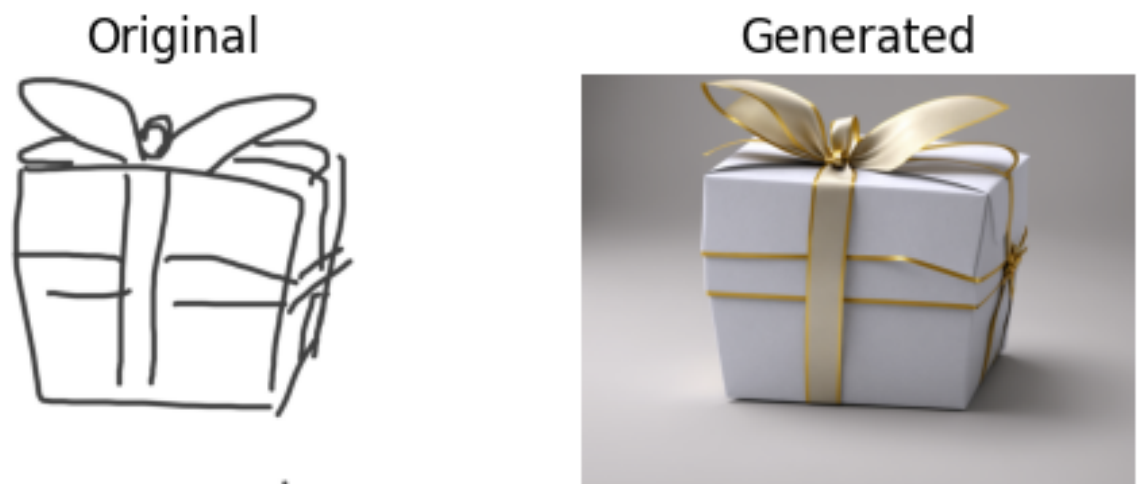
- *Контрольоване масштабування*: Під час збільшення кількості користувачів можна впровадити платні тарифи або ліміти для унеможливлення неконтрольованого зростання витрат.
- *Промпт-інжиніринг*: Уточнення промптів, щоб досягти потрібного результату з меншою кількістю повторних викликів.

## Швидкість роботи

- **Проблема:**
  - Аналіз зображення за допомогою “Analyze Images” триває в середньому 2–3 секунди.
  - Генерація реалістичного зображення зі скетчу (“Sketch to Image”) може займати до 25–30 секунд, що є досить тривалим для миттєвої взаємодії з користувачем.
- **Підхід до подолання:**
  - *Асинхронна обробка*: Застосування асинхронних викликів до API та реалізація механізму чергування завдань дає змогу не блокувати інтерфейс.
  - *Індикатор прогресу*: Користувач отримує повідомлення про поточний статус генерації зображення, що робить очікування менш проблемним.
  - *Кешування результатів*: Якщо користувач надсилає подібний ескіз чи повторно викликає генерацію, можна використати вже наявну інформацію (у випадках, коли це доречно).
  - *Оптимізація промптів*: Скорочення або уточнення описів може зменшити час генерації за рахунок мінімізації обсягу обчислень.

## Результати

Для порівняння було обрано наступне зображення:



Для скетчу було згенерований такий опис:

A realistically rendered, slightly angled view of a square gift box with a tied bow on top. The box is wrapped in paper with a vertical ribbon, creating

sections on the front. The box sits flat on a surface. The ribbon and bow should be realistically detailed with folds, the paper has a matte finish and is smooth. The lighting is soft and diffused, creating subtle shadows, giving the scene a calm and celebratory feel. The scene is a close-up, focusing on the details of the gift box. The colors of the bow and wrapping paper are vibrant and rich, creating a nice contrast.

## Отримані метрики

- **CLIP Similarity:**
  - *Original* – 25.28
  - *Generated* – 22.82
- **Object Match Score:**
  - *Original* – 0.00
  - *Generated* – 0.00
- **FID Score:**
  - *Original* – 0.21
  - *Generated* – 0.30
- **SSIM Metric:**
  - 0.78

Високе значення **SSIM** (0.78) вказує на доволі велику структурну схожість між оригінальним та згенерованим зображенням. Водночас, **CLIP Similarity** показує зменшення відповідності змісту (від 25.28 до 22.82), що може свідчити про дещо меншу релевантність згенерованого зображення відносно промпту/опису, порівняно з оригіналом. **FID** (Frechet Inception Distance) теж вищий у згенерованого зображення (0.30), але залишається на прийнятному рівні, враховуючи специфіку задачі (генерація зі скетчу).

## Порівняння з існуючими підходами

На ринку є чимало схожих інструментів для перетворення ескізів у цифрові зображення з використанням генеративного ШІ (наприклад, [OpenArt.ai](#), [Canva](#), [Adobe Firefly](#), [Freepik AI](#)).

1. **Зручність та спектр можливостей:** Більшість альтернатив пропонують більш широкий набір інструментів (зокрема стилізацію, додавання ефектів), однак можуть вимагати платної підписки або обмежувати кількість безплатних генерацій.
2. **Швидкість та вартість:** Подібні сервіси варіюються за часом обробки та вартістю доступу. Наш застосунок демонструє помірні показники швидкості (2–3 с для аналізу і 25–30 с для генерації) та чіткі тарифи за кожний запит, що робить витрати прозорими й передбачуваними.
3. **Контроль над процесом:** Власний проект дає можливість налаштовувати та інтегрувати додаткові модулі (наприклад, оцінку метрик) і безпосередньо керувати бекенд-частиною, чого може бракувати у готових SaaS-рішеннях.

Таким чином, попри наявні конкуренти, розроблене рішення вирізняється прозорістю ціноутворення, можливістю кастомізації й чітким контролем над генеративним процесом завдяки власному модулю метрик.

## Інтеграція

### API проекту GenAI Sketch-to-Image

Проект надає наступні API ендпоінти:

#### 1. Головна сторінка

- Метод: GET
- Ендпоінт: `/`
- Опис: Повертає головну сторінку застосунку (index.html).
- Використання: Відкрийте браузер за адресою `http://localhost:5050/`

#### 2. Обробка зображення

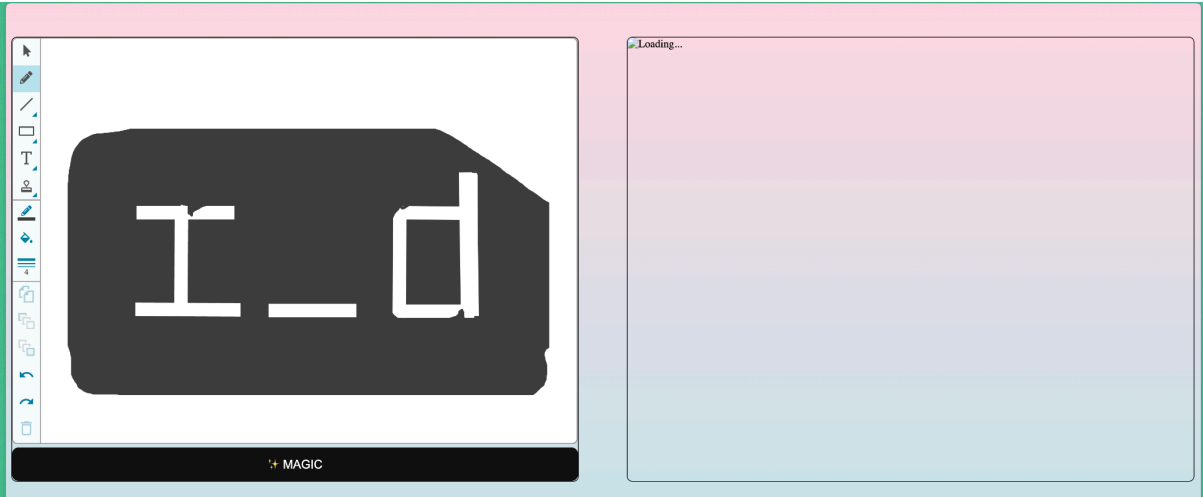
- Метод: POST
- Ендпоінт: `/magic`
- Опис: Обробляє завантажене зображення, генерує опис та конвертує ескіз у реалістичне зображення.
- Параметри запиту: JSON з полем `image` (base64-encoded зображення)
- Повертає: JSON з наступними полями:
  - `message`: Повідомлення про статус обробки
  - `filename`: Ім'я збереженого файлу
  - `description`: Згенерований опис зображення
  - `image`: URL згенерованого реалістичного зображення
  - `uuid`: Унікальний ідентифікатор обробленого зображення
- Коді відповіді:
  - 200: Успішна обробка
  - 400: Помилка обробки (з описом помилки)

#### 3. Статичні файли

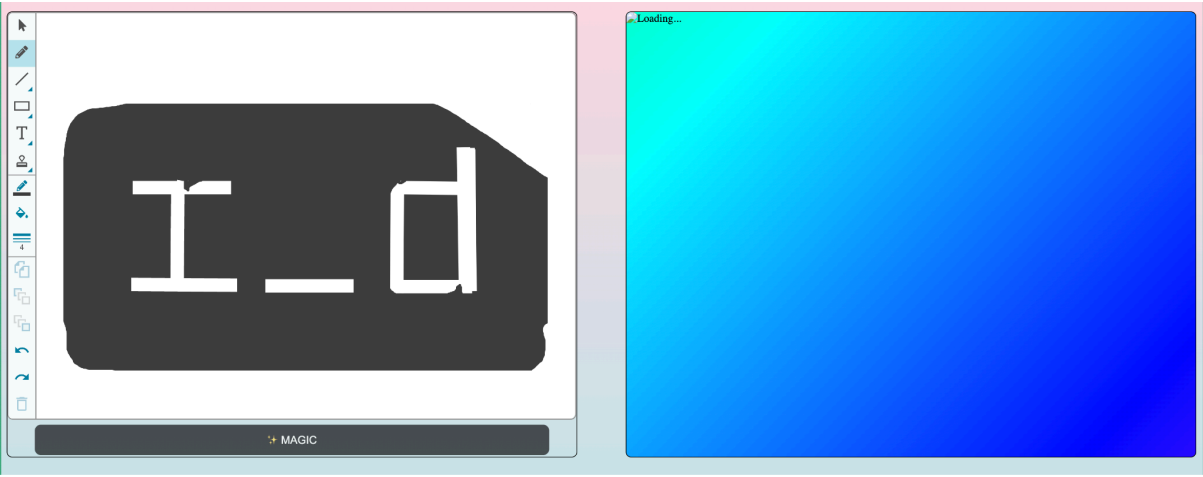
- Метод: GET
- Ендпоінт: `/assets/<path:filename>`
- Опис: Повертає статичні файли (CSS, JavaScript, зображення тощо).
- Використання: Автоматично використовується браузером для завантаження ресурсів.

### UI проекту GenAI Sketch-to-Image

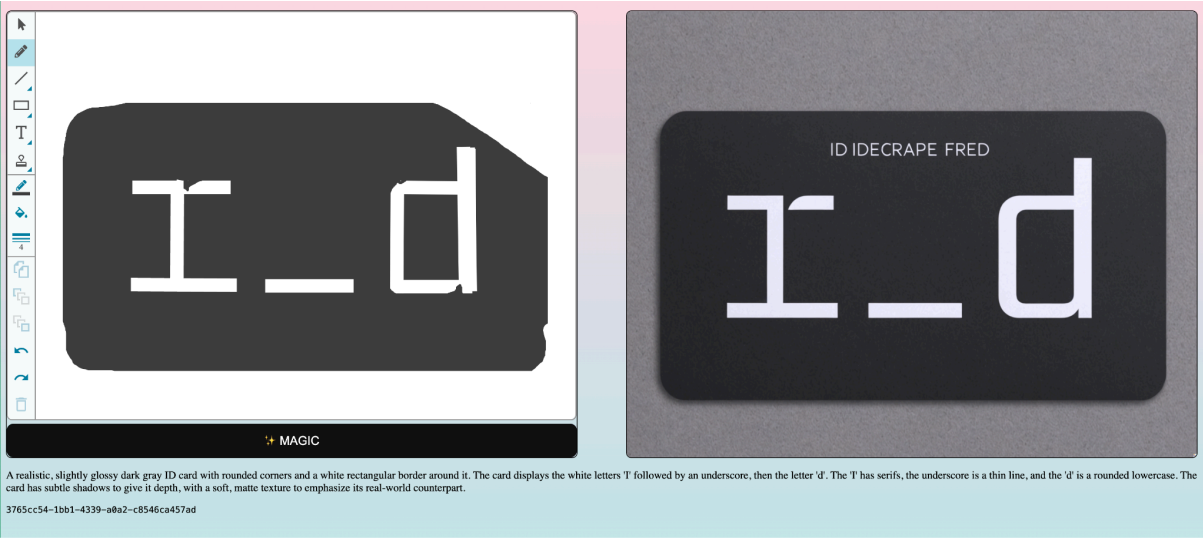
Процес малювання скетчу



Процес генерації малюнку (очікування результату)



Отримані результати: зображення та опис



# Висновки

## Оцінка досягнутих результатів

У рамках проекту вдалося створити повноцінну систему, здатну перетворювати скетчі користувача на реалістичні зображення із супровідними текстовими описами.

Результати метрик (CLIP Similarity, FID, SSIM) свідчать про прийнятний рівень якості згенерованих зображень з огляду на те, що в основі лежать готові моделі та API.

Інтегрований модуль метрик надає кількісну оцінку якості зображень, допомагаючи виявляти сильні та слабкі місця підходу. Також було збудовано зручний веб-інтерфейс, який забезпечує покрокову взаємодію користувача з системою.

## Майбутні можливості покращення

1. **Покращення опису:** Розширення промптів додатковим контекстом або залучення додаткових моделей для більш інформативних і доречних описів.
2. **Оптимізація швидкості:** Розглянути можливість організації обробки в асинхронних чергах або застосування локальних моделей, щоб зменшити час очікування (25–30 секунд) під час генерації зображень.
3. **Адаптація моделей:** Хоч проєкт наразі покладається на готові API, у перспективі можна розглянути fine-tuning моделей чи використання спеціалізованих моделей для скетч-ту-зображення, що потенційно покращить якість і знизить витрати.
4. **Пакетна обробка:** Додавши функцію одночасного завантаження кількох скетчів, користувачі зможуть ефективніше працювати над серією ідей чи концептів.
5. **Персоналізація:** Реалізувати систему авторизації, щоб кожен користувач мав власну історію завантажених скетчів і отриманих зображень, а також використовувати вподобання користувача для точнішого та релевантнішого результату.