

ВИКОРИСТАННЯ ГЕНЕРАТИВНОЇ МОДЕЛІ ШТУЧНОГО ІНТЕЛЕКТУ DALL-E У ВЕБДОДАТКАХ

Кирилов Д.І., Лебединський А.В.

Харківський національний автомобільно-дорожній університет, Харків

У сучасному світі штучний інтелект (ШІ) використовується в різних сферах, зокрема, в обробці зображень. Одним із найновітніших розробок є генеративна модель DALL-E, створена OpenAI, яка здатна генерувати унікальні зображення на основі текстового опису. Ця модель, базуючись на трансформерній архітектурі [1, 2], здатна створювати зображення на основі текстового опису, відкриваючи безмежні можливості для творчості та інновацій. У даній доповіді розглянемо можливості використання моделі DALL-E у вебдодатках та її потенційні переваги.

Хоча DALL-E залишається одним з найбільш відомих генераторів зображень штучного інтелекту, на ринку з'явилося кілька альтернатив, які також користуються популярністю. Midjourney і Stable Diffusion – два з найвідоміших з них, порівняння яких зроблено у [3]. Midjourney, розроблений незалежною лабораторією в Сан-Франциско, знаходиться у відкритому бета-тестуванні та може бути використаний через Discord. Цей інструмент славиться високою якістю та деталізацією, проте потребує оплати за створення зображень. Stable Diffusion, який має відкритий вихідний код, спочатку навчався на 2,3 мільярдах зображень і розроблений командою дослідників з різних установ, включаючи CompVis Group, StabilityAI та RunwayML. Популярність Stable Diffusion зростає, а його функціональність постійно розвивається завдяки активній спільноті користувачів.

DALL-E має різноманітні застосування, наприклад, створення власних зображень для соціальних мереж, ігор і додатків віртуальної реальності. Його можна використовувати в різних сферах, включаючи рекламу, маркетинг, графічний дизайн, відеоігри, а також дизайн продуктів. Наступник DALL-E, DALL-E 2, призначений для створення більш фотореалістичних зображень із вищою роздільною здатністю [4]. Інтерфейс DALL-E зображений на рис. 1.

Модель DALL-E AI пропонує розробникам низку можливостей API для використання розширених можливостей створення зображень. Ось основні можливості API для моделі DALL-E:

Інтеграція у продукти:

- DALL-E API дозволяє розробникам інтегрувати найсучасніші можливості створення зображень безпосередньо у свої продукти, дозволяючи їм створювати різноманітні та високоякісні зображення з текстових описів [4];
- розробники можуть вибрати саме ті функціональні можливості, які їм потрібні, не навантажуючи свій вебдодаток непотрібними бібліотеками і додатковими функціями [4].

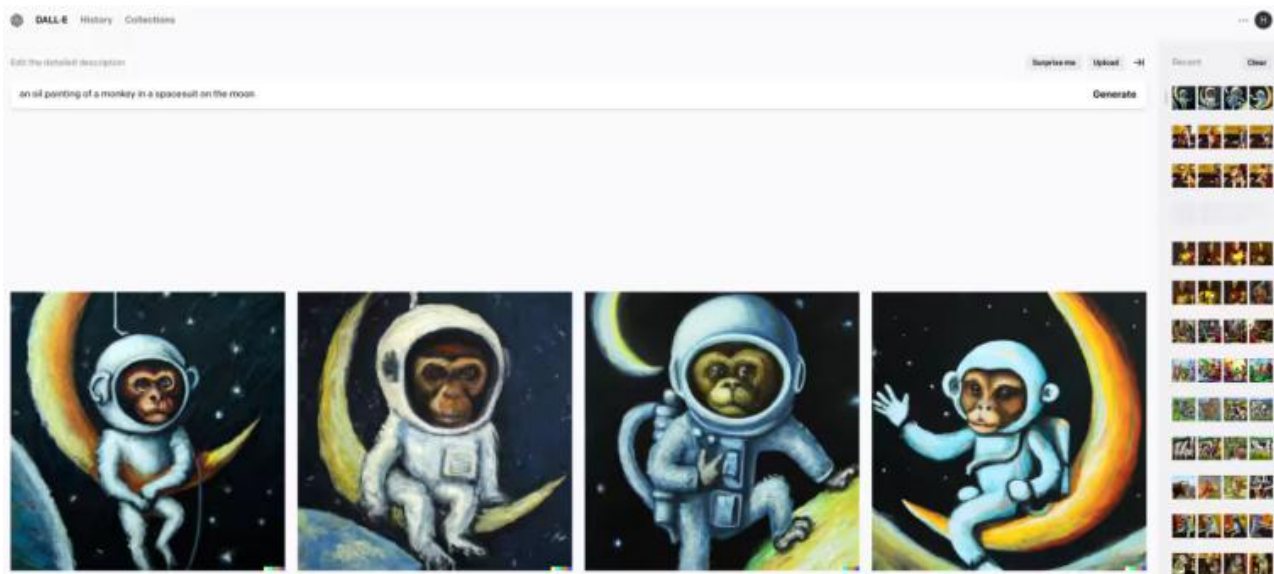


Рисунок 1 – Веб-інтерфейс моделі DALL-E

Модель «оплата по мірі використання»:

- DALL-E API працює на основі оплати за використання, що дозволяє користувачам отримувати доступ до його можливостей створення зображень на основі їх використання та потреб;

- інформацію про ціни на DALL-E API можна знайти на сторінці цін OpenAI, за запитом доступні варіанти знижок на великий обсяг.

Вихідне зображення та функції:

- DALL-E API може виводити зображення як URL-адреси або у форматі b64_json, забезпечуючи гнучкість доступу до згенерованих зображень і їх використання;

- за допомогою DALL-E 3 користувачі можуть скористатися такими функціями, як інтеграція тексту в зображення, що дозволяє бездоганно інтегрувати текст у візуальний контекст, а також можливість генерувати зображення в різних орієнтаціях, включаючи альбомний і портретний формати [5].

Саме наявність детальної документації для роботи з API дозволяє втілити у життя будь-які вебдодатки з використанням саме цієї функціональності. На рис. 2 показаний код функції generateImage, яка на вхід отримує аргументи вибору моделі, рядку запиту, роздільну здатність картинки, якість та кількість картинок.

Версії моделі та вдосконалення:

- API DALL-E підтримує різні версії моделі DALL-E, причому DALL-E 3 є останньою та найбільш вдосконаленою версією, яка пропонує покращене розуміння підказок і якість створених зображень [5];

- DALL-E 3 забезпечує зображення стандартної та високої чіткості з відповідною роздільною здатністю, задовольняючи потреби та вподобання різних користувачів [6].

```

1 from openai import OpenAI
2 client = OpenAI()
3
4 response = client.images.generate(
5     model="dall-e-3",
6     prompt="a white siamese cat",
7     size="1024x1024",
8     quality="standard",
9     n=1,
10 )
11
12 image_url = response.data[0].url

```

Рисунок 2 – Опис функції generateImage на NodeJS

Але з точки зору обмежень, створення зображень DALL-E не завжди може бути точним або відповідним, і можуть виникнути занепокоєння щодо порушення авторських прав, якщо створені зображення надто схожі на роботи, захищені авторським правом [7]. Модерація вмісту також є серйозною проблемою для запобігання створенню невідповідних, образливих або шкідливих зображень. Крім того, автоматизація створення контенту потенційно може витіснити робочі місця в таких сферах, як графічний дизайн, але також може відкрити нові ролі в нагляді та управлінні цими системами ШІ, але це вже інша тема для іншого дослідження.

Список використаних джерел

1. DALL-E 2. *OpenAI*. URL: <https://openai.com/dall-e-2> (date of access: 12.04.2024).
2. Учасники проєктів Вікімедіа. DALL-E – Вікіпедія. *Вікіпедія*. URL: <https://uk.wikipedia.org/wiki/DALL-E> (дата звернення: 12.04.2024).
3. Midjourney vs. Stable Diffusion: Which Should You Use? - viso.ai. *viso.ai*. URL: <https://viso.ai/deep-learning/midjourney-stable-diffusion/> (date of access: 12.04.2024).
4. Demir D. A Guide to Using the DALL·E 3 API: How to Use and Test it Online. *Apidog Blog*. URL: <https://apidog.com/blog/dalle-3-api/> (date of access: 12.04.2024).
5. API Reference. Getting started. *OpenAI*. URL: <https://platform.openai.com/docs/api-reference> (date of access: 12.04.2024).
6. Keita Z. A Comprehensive Guide to the DALL-E 3 API. *Learn Data Science and AI Online | DataCamp*. URL: <https://www.datacamp.com/tutorial/a-comprehensive-guide-to-the-dall-e-3-api> (date of access: 12.04.2024).
7. Awan A. A. What is DALL-E?. *Learn Data Science and AI Online | DataCamp*. URL: <https://www.datacamp.com/blog/what-is-dall-e> (date of access: 12.04.2024).