

ПЕТРИНА ДМИТРО

Івано-Франківський національний технічний університет нафти і газу

<https://orcid.org/0000-0003-0663-746X>e-mail: pepperiko@ukr.net

КОРНУТА ОЛЕНА

Івано-Франківський національний технічний університет нафти і газу

<https://orcid.org/0000-0002-0626-888X>e-mail: olena.kornuta@nung.edu.ua

ПОРІВНЯННЯ РІЗНИХ ВЕРСІЙ ДОДАТКУ MIDJOURNEY ТА ЙОГО ВИКОРИСТАННЯ В РІЗНИХ СФЕРАХ СТВОРЕННЯ ВІЗУАЛЬНОГО КОНТЕНТУ

В статті порівнюються якісні параметри візуального контенту, створеного на базі штучного інтелекту програмою Midjourney з використанням різних її версій. Проаналізовані параметри написання пояснень (prompts), які впливатимуть на остаточну якість згенерованого зображення та можливе використання програми для різних сфер діяльності людини.

Ключові слова: штучний інтелект, Midjourney, Leonardo, prompt, нейромережа.

PETRYNA DMYTRO, KORNUTA OLENA

Ivano-Frankivsk National Technical University of Oil and Gas

COMPARISON OF DIFFERENT VERSIONS OF THE MIDJOURNEY AND ITS USE IN VARIOS SPHERES OF CREATION VISUAL CONTENT

To date, the use of applications for the generation of illustrative material with the help of artificial intelligence (AI) is one of the most progressive for creating content in the field of visual occupations (designers, architects, artists), for marketers, students and ordinary people. The main reason for using neural networks is to save time and create inspiring examples in any field of human activity. Currently, there are more than 20 independent programs that generate visual content, and many companies such as Adobe and Canva use neural network tools. The use of artificial intelligence is irreversible and requires practice and some experience in its work with the creation of visual content. The development and updating of the main programs in this direction, such as Midjourney and Leonardo, is constant and needs to be studied. The photo-realism and image detailing of the latest versions of Midjourney allows you to create visual content that can be used to generate new ideas and use it for creation quality visual content for advertising. The article compares the quality parameters of the visual content created on the basis of artificial intelligence by the Midjourney program using its different versions. The updates from the initial version to the latest V6.0 were phased in less than a year and a half. Certain deviations in the images are considered, which leads to the impossibility of their further use. Analyzed the parameters of writing explanations (prompts), which will affect the final quality of the generated image, the possibility of writing complex prompts with one and several images, as well as advanced prompts. The possible use of neural networks that work with text for targeted writing of appropriate prompts for better automation of the generation process is considered. The article considers the possible use of the Midjourney program for various spheres of human activity that use various images, and also raises the question of the role of a person as the main creator and generator of creative ideas.

Keywords: artificial intelligence, Midjourney, Leonardo, prompt, neural network.

Постановка проблеми

На теперішній час існує безліч додатків, які використовують штучний інтелект (ШІ) для генерації, обробки та редагування візуального контенту. Провідні компанії такі як Adobe та Canva використовують можливості нейромереж для пришвидшення та автоматизації процесів обробки. Через простоту та доступність у використанні ШІ одними з перших та постійних користувачів таких додатків є дизайнери, архітектори, візуалізатори, маркетологи, студенти. Сьогодні неможливо уявити процес навчання без генерування нових ідей, адже основна перевага такого інструментарію – це швидкість отримання інформації та її візуалізації. Постійне вдосконалення цих додатків призводить до отримання більш якісного результату.

Однією з провідних нейромереж для генерації візуального контенту є додаток Midjourney, який одним з перших почав використовувати ШІ. Через велику популярність додаток є повністю платним, проте для творчого процесу дизайну (графічного, UI/UX дизайну, веб дизайну, 3D моделювання), архітектури, пошиття одягу, спеціальностей художнього спрямування він створює якісний візуальний об'єкт і дозволяє генерувати та змінювати вже готові результати. Швидкий розвиток програми за останні півтори роки суттєво змінив у якісному плані генеровані зображення. Постійні зміни налаштувань та використання новітніх версій потребують певного вивчення та практичного удосконалення.

Аналіз останніх джерел

На даний час існує безліч нейромереж, які застосовують штучний інтелект для створення візуального контенту, проте є кілька топових, які зарекомендували себе лідерами у створенні генеративного зображення [1, 2, 3]. У грудні 2020 року кабінет міністрів України видав розпорядження «Про схвалення Концепції розвитку штучного інтелекту в Україні» [4], де дав короткі рекомендації щодо подальшого розвитку та використання ШІ у різних сферах людської діяльності та навчання. Проте швидкий розвиток нейромереж потребує постійного моніторингу, навчання та практики. Постійні зміни та вдосконалення не дають можливості ґрунтовного дослідження і тому основними джерелами є інструкції для використання тих чи інших програм, а також різноманітні навчальні відеоролики на youtube [5].

Метою роботи є порівняння зображень, згенерованих попередніми версіями додатку Midjourney з останньою версією V 6.0 та програмою Leonardo AI, розглянути основні переваги та недоліки, які виникають в процесі роботи програми, описати процес написання промптів для створення нових зображень, а також окреслити основні області застосування візуальних нейромереж та спробувати визначити роль людини як творця нового контенту при роботі з нейромережами.

Виклад основного матеріалу

Midjourney — незалежна дослідницька лабораторія та назва її власної програми штучного інтелекту, яка створює цифрові зображення з описів природною мовою за допомогою «підказок» (від англ. prompt). З 12 липня 2022 року програма перебуває у відкритому бета-тестуванні. Лабораторія фінансується самостійно, її діяльність зосереджена на дизайні, людській інфраструктурі та штучному інтелекті [6]. За цей короткий час існування програма розвивається і самонавчається. На початку свого існування програма працювала у версії 1 (V 1.0), на сьогодні проходить тестування 6-та версія (V 6.0 ALFA). Зараз є можливість спробувати всі версії цієї лабораторії, перемикаючись між ними (рис 1). Проте, варто зазначити, що при генеруванні зображень у більш старіших версіях збільшується час під'єднання.

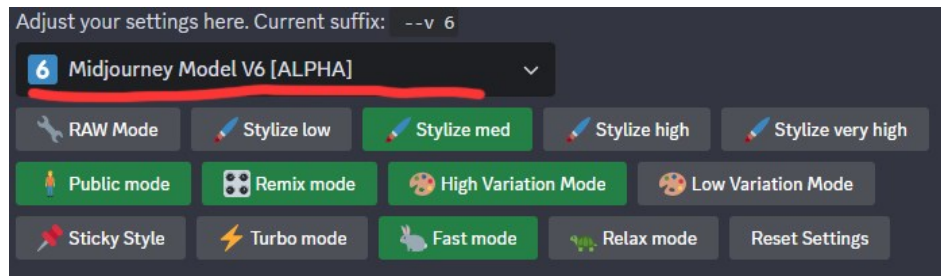


Рис. 1. Зміна версії моделі Midjourney

Програма використовує власну стилістику, проте користувач може за допомогою налаштувань змінити або взагалі не враховувати вподобання нейромережі (Raw Mode). Існує безліч налаштувань (settings) які дозволяють керувати процесом генерації і на сайті лабораторії є детальне описання кожної функції [7].

Нашою основною задачею було порівняння зображень, створених різними версіями Midjourney та оцінка переваг чи недоліків останніх версій. Оскільки кожне з налаштувань може ввести певні відхилення в порівняльний аналіз різних версій, то для процесу генерації використовували однакові налаштування, що діють за замовчуванням.

Для початку генерування зображень у нейромережі необхідно оплатити підписку. До недавнього часу існувала безкоштовна версія, яка дозволяла створювати обмежену кількість зображень, проте з ростом популярності програми у користувачів різного рівня підготовки та зацікавленості розробники ввели обов'язкову плату і на даний час найдешевша версія підписки складає 9\$.

Основною генерацією займається чат бот discord.com, який необхідно встановити на комп'ютері, або використовувати онлайн версію. Мінімальна підписка дає можливість генерувати зображення у спільному чаті, що може призвести до певних незручностей через велику кількість відвідувачів, проте дозволяє навчатися у більш досвідчених користувачів (рис. 2).

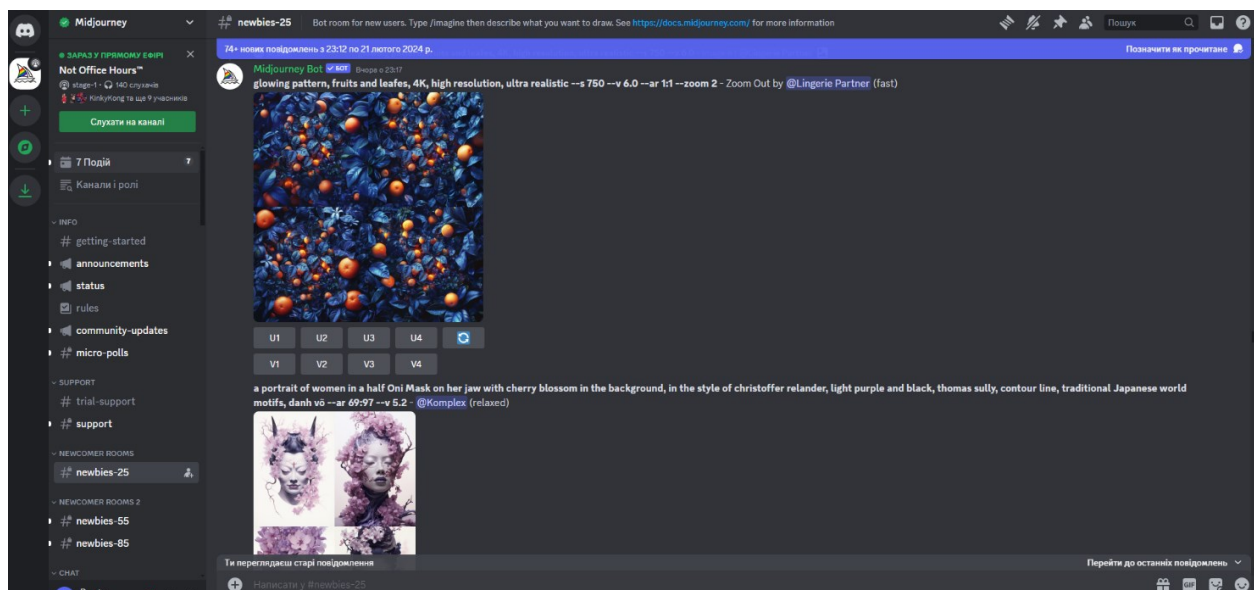
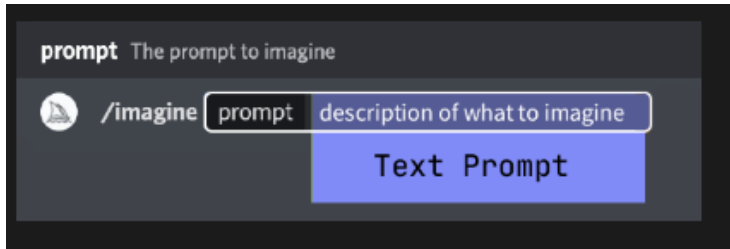
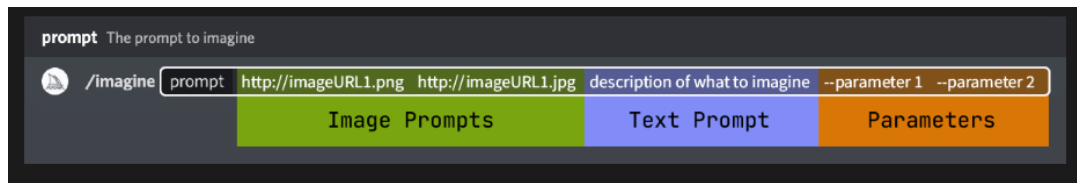


Рис. 2. Інтерфейс чат боту discord.com

Основою створення будь якого зображення у неймережі є промт (prompt). Промт — це коротка текстова фраза, яку Midjourney Bot інтерпретує для створення зображення. Бот Midjourney розбиває слова та фрази в промті на менші фрагменти, які називаються маркерами, які порівнюються з даними навчання, а потім використовуються для створення зображення. Добре складений промт може допомогти створити унікальні та захоплюючі зображення. Розрізняють як прості (рис. 3а) так і комбіновані промти (рис. 3б). Комбіновані промти окрім основної (текстової) частини містять посилання на зображення, що завантажене в чат боті та параметричні доповнення, які суттєво впливають на створювання кінцевого зображення.



а



б

Рис. 3. Види написання промтів: а) простий (текстовий) промт; б) комбінований промт

Від точності та лаконічності промтів, конкретизованого пояснення кожної деталі, стилю в якому буде сформоване зображення буде залежати остаточний результат генерації. Деякі кваліфіковані та досвідчені користувачі такі як Джонас Петерсон претворили звичайну і типову генерацію на справжні мистецькі твори [8].

Як вказано раніше, параметричні дані (parameters) будуть суттєво впливати на остаточний результат і вмиле їх використання дозволить отримувати більш точні, якісніші та при потребі дивні та непередбачувані зображення. Параметри вписуються в текстові промти за допомогою символів, літер та цифрових значень. В таблиці 1 вказані деякі з цих параметрів.

Таблиця 1

Позначення та розшифрування певних параметрів промтів

Символьне позначення	Пояснення
--ar 2:3	Параметр змінює співвідношення сторін згідно вказаних числових співвідношень. За замовчуванням створюється співвідношення зображення 1:1
--chaos 20	Параметр в залежності від цифрового значення (0-100) буде створювати дивні та непередбачувані зображення. За замовчуванням значення цього параметру дорівнює 0
--no	Негативний промт буде запобігати візуалізації предметів, вказаних після символічного позначення
--s 500	Використання стилізації Midjourney (0-1000). За замовчуванням значення цього параметру дорівнює 100
--tile	Дає можливість створювати безшовні текстири
--w 300	Параметр «дивакуватий» (0-3000) дозволяє створювати випадкову стилізацію та експериментувати з нею. За замовчуванням значення цього параметру дорівнює 0

Для порівняння різних версій Midjourney використовувався простий текстовий промт: women in the blooming Japanese garden --ar 2:3 (жінка в квітучому саду в японському стилі, співвідношення сторін 2:3). Основні налаштування чат боту та налаштування параметрів виставлялися за замовчуванням. Також для порівняння створювалася генерація в додатку Leonardo AI з тим самим промтом. Результати генерації зображень наведені на рис. 4

Згідно візуального порівняння згенерованих картинок видно, що за допомогою останньої версії Midjourney V6.0 (рис. 4, д) створена візуальна ілюстрація характеризується найбільшою чіткістю, крайовою різкістю, більше нагадує фотографію, ніж графіку чи рисунок. Версія V5.2 (рис. 4, г) також створила непоганий результат, який симулює фотографічне зображення з невеликою глибиною різкості, проте дещо уступає реалістичності нової версії. Попередні версії показують різний результат, проте він більше нагадує комп'ютерну ілюстрацію або рисунок аквареллю.



а



б



в



г



д



е

Рис. 4. Результати генерації промту різними версіями програми Midjourney та Leonardo AI:
а) V4.0; б) V5.0; в) V5.1; г) V5.2; д) V6.0; е) Leonardo AI

Генерація в Leonardo AI є значно простішим варіантом генерації за однаковим промтом, який має чітко виражений ілюстративний характер. Причиною цього може бути використання безкоштовної версії додатку та застосування стилістики, притаманної для Leonardo. Проте Midjourney в різних версіях показує кращу девіацію при однакових промтах та можливість більш розширеного керування налаштуваннями та параметризацією.

Якщо говорити про недоліки візуальної лабораторії Midjourney, то необхідно сказати про дивну деталізацію та стилістику при різних налаштуваннях промтів. Так, наприклад, для попереднього промту з додаванням параметру стилістики --s 500 отримані певні артефакти (позначені червоною обводкою) на зображенні (рис. 5).



Рис. 5. Прояв артефактів на певних зображеннях

При генеруванні зображень людей в повний зріст програма досить часто дає функціональні збої, що призводить до неправильної або дивної анатомічної структури людей у вигляді неправильної форми кінцівок, низької деталізації або певних складних конструкцій (рис. 6).



Рис. 6. Анатомічна невідповідність людей

Певні артефакти усуваються у додатковому редагуванні в програмі Adobe Photoshop, при повторній генерації промту, або корекції вже існуючої версії за допомогою команди Vary(region) і подальшою генерацією, незважаючи на те, що такі шляхи можуть призвести до втрати часу. Проте по факту швидкого оновлення більш старших версій можна сказати, що програма самовдосконалюється і слід очікувати виправлення і усунення невідповідностей у анатомії людей та конструкції складних об'єктів.

Що стосується правової власності створених об'єктів, то згідно політики лабораторії Midjourney будь-які зображення, що згенеровані платною версією, є повною власністю користувача, який здійснював генерацію і таким чином, використання ілюстрацій в будь-яких проявах, а також комерційне використання є повністю легальне і не потребує додаткової конфірмації [9].

Висновки

Нейромережі з можливістю генерування зображень є сучасним інструментом, який вже змінює та полегшує роботу людини в різноманітних творчих та прикладних аспектах діяльності та навчання. Основна їх перевага – це економія часу та можливість створювати нові цікаві ідеї, базуючись на попередньому досвіді. Багато ілюстративних компаній інтегрують можливості ШІ в різноманітні інструменти та додатки.

Незалежна дослідницька лабораторія Midjourney є флагманом в сфері генерації зображень. Достатньо часті оновлення програми показують її швидкий і постійний розвиток, а її велика популярність дозволяє говорити про широке застосування та незворотність процесу інтеграції ШІ у різні види людської діяльності. Остання версія Midjourney V6.0 дає можливість отримати якісне зображення фотографічної якості у порівнянні з попередніми версіями програми та найближчими конкурентами. Проте слід зазначити, що ґрунтовне вивчення всіх можливих програм з генерації потребує значного часу та фінансування. Звичайно, існують певні недоліки та неточності генерування, що виникають в роботі, проте швидке оновлення та створення нових версій дає надію на покращення цієї ситуації.

Потрібно зазначити про важливість інтегрування та вивчення програм з ШІ в освітню діяльність. Мистецькі, дизайнерські, технічні спеціальності, маркетинг та медицина – напевно важко назвати сферу, де б візуалізація процесу на базі нейромереж не прискорила та покращила процес навчання. Тому навчання спеціалістів, які володіють відповідними навиками – одна з задач сучасної освіти України.

Література

1. A. M. Radhakrishnan, "Is midjourney-ai a new anti-hero of architectural imagery and creativity?", GSI, vol. 11, №(1), pp. 94-104, 2023
2. Капанжи С. О. Дослідження інструментів штучного інтелекту в професійній діяльності викладача. The 5th International scientific and practical conference "Innovative development of science, technology and education" (February 15-17, 2024) Perfect Publishing, Vancouver, Canada. 2024. p. 173-181.
3. A. Jaruga-Rozdolska, "Artificial intelligence as part of future practices in the architect's work: MidJourney generative tool as part of a process of creating an architectural form", Architectus, vol. 71, № 3, pp. 95- 104, 2022. <https://doi.org/10.37190/arc220310>
4. КАБІНЕТ МІНІСТРІВ УКРАЇНИ РОЗПОРЯДЖЕННЯ від 2 грудня 2020 р. № 1556-р «Про схвалення Концепції розвитку штучного інтелекту в Україні». URL: <https://zakon.rada.gov.ua/laws/show/1556-2020-%D1%80#Text>
5. Advanced Midjourney V6 Guide (Pushing Boundaries of Lifelike Cinematic AI Photography).

- URL: https://youtu.be/Zl_4V0ks7CE?si=m_E7HF9c2rbzpjES
6. Midjourney. URL: <https://www.midjourney.com/home>
7. Midjourney bot. URL: <https://docs.midjourney.com/v1/docs>
8. Jonaspeteronai. URL: <https://www.instagram.com/jonaspeteronai?igsh=MWV1anpoYWJnemJvMw==>
9. Midjourney Privacy Policy. URL: <https://docs.midjourney.com/docs/privacy-policy>

References

1. A. M. Radhakrishnan, "Is midjourney-ai a new anti-hero of architectural imagery and creativity?", GSI, vol. 11, №(1), pp. 94-104, 2023
2. Kapanzhy S. O. Doslidzhennia instrumentiv shtuchnoho intelektu v profesiinii diialnosti vykladacha. The 5th International scientific and practical conference "Innovative development of science, technology and education" (February 15-17, 2024) Perfect Publishing, Vancouver, Canada. 2024. p. 173-181.
- 3 A. Jaruga-Rozdolska, "Artificial intelligence as part of future practices in the architect's work: MidJourney generative tool as part of a process of creating an architectural form", Architectus, vol. 71, № 3, pp. 95- 104, 2022. <https://doi.org/10.37190/arc220310>
4. KABINET MINISTRIV UKRAINY ROZPORIADZHENNIA vid 2 hrudnia 2020 r. № 1556-r «Pro skhvalennia Kontseptsii rozvytku shtuchnoho intelektu v Ukraini». URL: <https://zakon.rada.gov.ua/laws/show/1556-2020-%D1%80#Text>
5. Advanced Midjourney V6 Guide (Pushing Boundaries of Lifelike Cinematic AI Photography). URL: https://youtu.be/Zl_4V0ks7CE?si=m_E7HF9c2rbzpjES
6. Midjourney. URL: <https://www.midjourney.com/home>
7. Midjourney bot. URL: <https://docs.midjourney.com/v1/docs>
8. Jonaspeteronai. URL: <https://www.instagram.com/jonaspeteronai?igsh=MWV1anpoYWJnemJvMw==>
9. Midjourney Privacy Policy. URL: <https://docs.midjourney.com/docs/privacy-policy>