УДК 004.8+655.262+003.21

ПІДГОТОВКА ІЛЮСТРАЦІЙ ДЛЯ ІНКЛЮЗИВНОЇ ЛІТЕРАТУРИ ЗА ДОПОМОГОЮ МОДЕЛЕЙ ШТУЧНОГО ІНТЕЛЕКТУ СИНТЕЗУ ЗОБРАЖЕННЯ З ТЕКСТУ

Є. А. Джуринський, В. З. Маїк

Українська академія друкарства,
вул. Під Голоском, 19, Львів, 79020, Україна

Значною проблемою в друкованій інклюзивній літературі є підготовка ілюстрацій, які мають передавати інформацію читачеві, що заклав автор, у графічний спосіб. Зважаючи на те, що цільовою аудиторією інклюзивної літератури є люди із порушеннями зору, варто пам'ятати про дотримання вимог до таких ілюстрацій, враховуючи як технічні обмеження засобів друку, так і особливості тактильної та нервової системи людини. Однією з нагальних проблем у сфері випукло-тактильних ілюстрацій є необхідність витрачати велику кількість часу на підготовку навіть однієї ілюстрації. Ця проблема передусім пов'язана із дефіцитом компетентних кадрів із освітою у сфері образотворчого мистецтва, що мають навички підготовки зображення як випукло-тактильної ілюстрації в інклюзивній літературі. Крім того, підготовка такого зображення відбувається ймовірніше в інтуїтивний спосіб, який часто визначається конкретною редакцією або друкарнею, що заважає узгодженому розвитку ілюстрації в інклюзивній літературі. Розглянуто засоби штучного інтелекту, що здатні вирішити наведені проблеми. Серед таких засобів були досліджені моделі синтезу зображення з тексту: Miourney, Stable Diffusion, DALL·E 2. Ці моделі є потужним інструментом, що здатен вирішувати великий спектр задач, забезпечуючи високий рівень унікальності та варіативності результатів. Головною перевагою таких засобів є автоматизація процесу підготовки ілюстрації. Насамперед автоматизація цього процесу може вирішити одну з нагальних проблем цієї сфери — дефіцит компетентних кадрів. Крім того, великою перевагою автоматизованого підходу є значне збереження часу — ілюстрація, яка раніше могла готуватися годинами або днями у традиційний спосіб, за допомогою штучного інтелекту ця задача може вирішуватися за дуже короткий час (секунди або хвилини). Проведено експериментальні дослідження для визначення можливостей різних моделей штучного інтелекту щодо перетворення тексту в зображення (ілюстрації) для інклюзивної літератури, дотримуючись при цьому головних принципів та вимог до процесу підготовки таких ілюстрацій. Після проведення таких експериментів було виявлено, що, незважаючи на те, що наведені засоби не здатні повністю готувати зображення як ілюстрації для інклюзивної літератури, бо не задовольняють вимоги до таких ілюстрацій повною мірою, проте такий метод має потенціал і може бути використаним у вирішенні цієї проблеми, взявши за основу принцип роботи таких рішень.

Ключові слова: люди з проблемами зору, *друк, штучний інтелект, синтез зображення, перетворення тексту у зображення, Midjourney, Stable Diffusion, DALL·E 2, інформаційна технологія, інклюзивна технологія, інклюзивна література, тактильна література, зображення, тактильна ілюстрація, вимоги до ілюстрації, ергономіка.*

Постановка проблеми. Процес розробки та підготовки ілюстрацій для інклюзивної літератури передбачає чималу кількість обмежень, яких має дотримуватися розробник таких ілюстрацій (або дизайнер) задля досягнення головної мети — передача графічної інформації читачеві із порушеннями зору без втрати найважливішої інформації. Як наслідок, цей процес потребує відповідних компетенцій від розробника, зокрема, пов'язаних із технічним виконанням випукло-тактильних ілюстрацій [1].

Оскільки галузь ілюстрації в інклюзивній літературі є досить вузькою і не має широкого розповсюдження серед спеціалістів образотворчого мистецтва, пошук та навчання дизайнера для розробки випукло-тактильних ілюстрацій є нетривіальною задачею.

Крім того, у редакцій інклюзивної літератури переважно немає чітко визначеного бачення того, як має виглядати результуюче зображення; а вимоги до ілюстрації одного і того ж предмета можуть відрізнятися, залежачи від цільової аудиторії, найважливішої інформації, яку хоче донести автор тощо.

Також відсутність чітких критеріїв до випукло-тактильних зображень та помірний розвиток галузі у відповідному професійному середовищі робить процес підготовки ілюстрацій в інклюзивній літературі важким, довгим і неузгодженим.

Аналіз останніх досліджень та публікацій. Останнім часом чималого розвитку набувають засоби штучного інтелекту, які дають змогу генерувати зображення, спираючись на текстову підказку користувача (перетворення тексту в зображення). Серед сучасних моделей до таких засобів належать Midjourney [2], Stable Diffusion [3] та DALL·E 2 [4].

Наведені моделі є так званими дифузними моделями [5, 6], які побудовані з ієрархії шумопоглинаючих автокодерів (від англ. — autoencoder, або AE). Вони показали, що досягають вражаючих результатів у синтезі зображень, бо дають змогу моделювати мультимодальні розподіли, що добре підходить для синтезу як фотореалістичних, так і фіктивних зображень. Такі моделі тренуються на мільйонах пар текст-зображень і складаються з мільярдів параметрів, що навчаються впродовж довгого часу, а процес розробки таких моделей потребує великого обчислювального ресурсу, який забезпечується суперкомп'ютерами, які не є доступними для пересічного користувача.

Варто зазначити, що дифузні моделі працюють скоріше в абстрактний спосіб, але не імперативний [7]. Це означає, що алгоритм синтезує зображення, спираючись на широку базу опрацьованих пар тексту-зображення, що призводить скоріше до утворення асоціативних зв'язків (проміжні результати обчислюються, спираючись на розподіл даних, що інтуїтивно можна описати як асоціативні кластери, подібно

до того, як працює людський мозок), ніж до чіткого та однозначного відображення слова в елемент зображення.

Однією з найпопулярніших моделей на сьогодні є Midjourney. Вона, як і інші наведені моделі у цій статті, є дуже складною та масштабною (у контексті об'єму тренувального набору даних та параметрів тренування) моделлю, що дає змогу їй виконувати задачі широкого спектра. За її допомогою користувач, що не обов'язково має володіти високим рівнем компетенції в образотворчому мистецтві, має можливість створювати деталізовані зображення вражаючої якості та унікальності, надавши лише текстову підказку із описом того, що користувач хоче побачити у результаті.

Мета статті — визначити можливості моделей штучного інтелекту перетворювати текст в зображення, створювати (або синтезувати) ілюстрації для інклюзивної літератури, дотримуючись при цьому головних принципів та вимог до процесу підготовки таких ілюстрацій.

Виклад основного матеріалу дослідження. Часто ілюстрація в інклюзивній літературі супроводжується текстом, що описує зображення. Такий опис може бути як художнім, так і суто технічним, тобто таким, який зосереджується на тому, як саме зображений предмет ілюстрації: з чого предмет складається, як елементи розташовані та залежать один від одного тощо.

З урахуванням того, що дифузні моделі синтезу зображення виконують задачі широкого спектра і здатні обробляти текст як вхідний параметр, поставимо експеримент і з'ясуємо, наскільки їх результати можуть бути застосовані у галузі інклюзивної літератури, не забуваючи про те, що випукло-тактильні ілюстрації мають відповідати обмеженням та вимогам таких ілюстрацій.

Розглянемо синтез зображення для інклюзивної літератури на прикладі роботи моделей Midjourney, Stable Diffusion та DALL·E 2. Оскільки ці моделі перетворюють текст в зображення, ми маємо визначити запит, який буде описувати зображення, яке ми хочемо отримати у кінцевому підсумку. Всі наведені моделі здатні оброблювати природну мову (від англ. — natural language processing), тому текстовий запит може містити більш детальний опис очікуваного зображення.

Запит (або текстова підказка), який ми будемо надавати моделям, буде містити такий текст: «*a primitive, flat, unfilled 2d illustration of a tree with no extra details with a contour composed of gapped dots*» (у перекладі з англ. — «*примітивна, плоска, незаповнена двовимірна ілюстрація дерева без зайвих деталей із контуром, що складається з точок із проміжком*»). За допомогою такої підказки ми очікуємо отримати примітивне зображення дерева, яке можна буде використовувати в інклюзивній літературі. Результати роботи моделей на такий запит наведені на рис. 1.

Як ми можемо бачити, дивлячись на рисунок, отриманий набір зображень є примітивним та спрощеним (хоча і в різному ступені, якщо роздивлятись кожне зображення окремо), крім того, об'єкти розташовані у фронтальний спосіб (або анфас), що означає, що спостерігач може бачити всі кінцівки дерева, не втрачаючи найважливішої інформації.

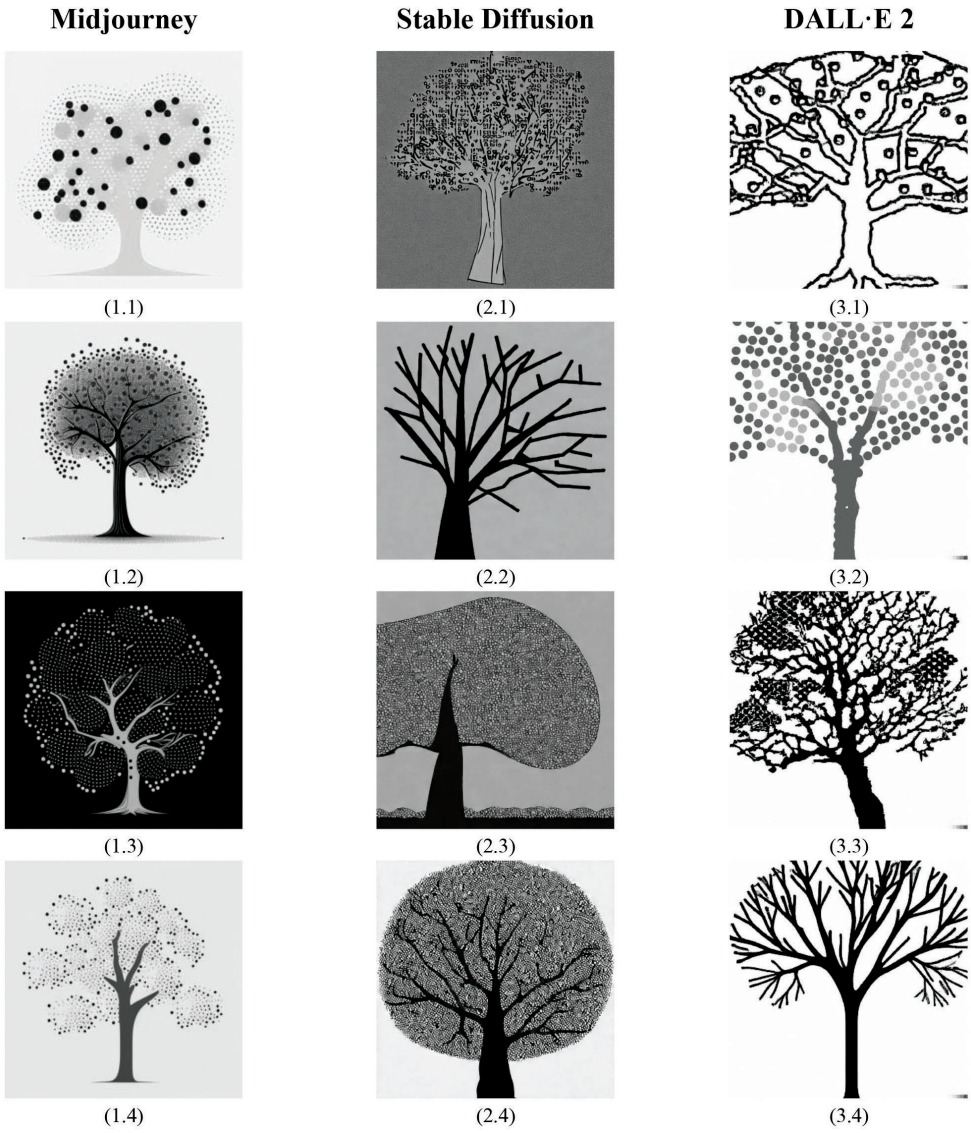


Рис. 1. Приклади зображення дерева для інклюзивної літератури, створеного за допомогою моделей синтезу зображення з тексту

Проте зображення містять певну низку невідповідностей критеріям ілюстрації в інклюзивній літературі, і тому вони потребують доопрацювання з боку компетентного ілюстратора:

Велика кількість зайвих деталей, що будуть відігравати роль «інформаційного шуму» для читача із порушеннями зору;

Перекривання одного елемента об’єкта зображення іншим елементом, що призведе до втрати інформації читачем із порушеннями зору;

Отримані об’єкти зображення мають химерну, асиметричну форму, що позитивно впливає на художні якості зображення, але негативно на якість сприйняття інформації читачем із порушеннями зору;

Відсутність спрощення та примітивізації, що ускладнює читачеві розпізнати найголовніші елементи, що характеризують об’єкт зображення.

Порівнюючи синтезовані зображення різних моделей на один і той самий текстовий запит, можемо зазначити, що кожна з цих моделей (незважаючи на схожу архітектуру і принцип роботи) має унікальний спосіб відображення елементів зображення: Midjourney синтезує зображення у більш художньому стилі, Stable Diffusion синтезує більш абстрактні зображення, а DALL·E 2 — графічне.

Потрібно зазначити, що серед наведених моделей, найкраще для ілюстрацій в інклюзивній літературі підходить модель DALL·E 2, оскільки ця модель краще розпізнає структурні графічні елементи, що є корисно для випукло-тактильних ілюстрацій.

Не можна оминути той факт, що результати роботи моделей у вигляді зображення на вихідну текстову підказку не є детермінованими та ідемпотентними, що означає, що кожний новий запит, хоча і з однаковим змістом запиту, буде давати новий результат. На рис. 2 зображений результат роботи моделі Midjourney із такою ж самою текстовою підказкою, як і для рис. 1, проте виконаним окремим запитом.

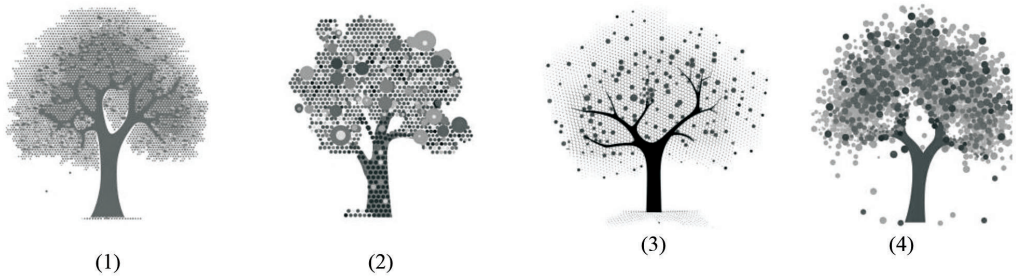


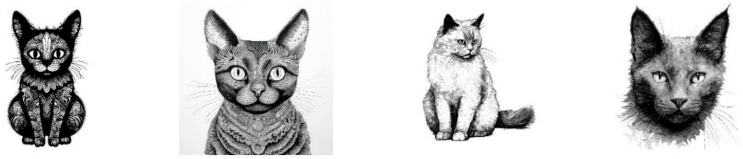
Рис. 2. Приклад зображення дерева для інклюзивної літератури, синтезованого за допомогою моделі Midjourney

Як можна побачити, результат на той самий текстовий запит відрізняється від того, що ми бачили попередньо. Така особливість пов’язана із стохастичною природою наведених моделей і, як наслідок, вона може негативно впливати на роботу під час підготовки ілюстрацій, оскільки в цьому випадку користувач має надавати моделі інший (тобто видозмінений) текстовий запит, намагаючись досягти бажаного результату, навіть якщо цей же самий запит попередньо працював у задовільний спосіб. З іншого боку, користувачу забезпечується висока варіативність та унікальність бажаного зображення, бо він зможе обирати найкраще зображення з набору запропонованих. Проте, зважаючи на те, що такі моделі за визначенням є стохастичні, від цієї особливості, яка може відігравати як позитивну, так і негативну роль, позбутися неможливо.

Також зазначимо, що всі наведені моделі є у відкритому доступі і їх навчання відбувається у постійний спосіб в режимі реального часу, спираючись на запити звичайних користувачів. Цей факт може відігравати негативну роль для використання наведених моделей в інклюзивній ілюстрації, оскільки використання алгоритму в широкому спектрі задач диверсифікує розподіли параметрів моделі, що якраз можуть бути потрібними для синтезу випукло-тактильних ілюстрацій.

Таблиця 1

Синтезовані зображення за допомогою моделі Midjourney

Текстова підказка	*coloring page for kids, black outline, white background, a tree, cartoon style, very low detail, no shading* (у перекладі з англ. — «*розмальовка для дітей, чорний контур, білий фон, дерево, мультиплікаційний стиль, дуже низька деталізація, без затінення*»)
Результат	
Текстова підказка	*black outline, white background, a cat without fur, primitive, contour only, very low detail, no shading* (у перекладі з англ. — «*чорний контур, білий фон, кіт без шерсті, примітивний, лише контур, дуже низька деталізація, без затінення*»)
Результат	
Текстова підказка	*white background, a simple black silhouette of a cat, side look* (у перекладі з англ. — «*білий фон, простий чорний силует кота, погляд збоку*»)
Результат	

Не можна не згадати той факт, що інша вхідна текстова підказка буде давати зовсім інший результат синтезованого зображення, тому так важливо у правильний

спосіб надавати текстовий запит. У табл. 1 наведені приклади синтезованого зображення із різними текстовими підказками, що синтезовано за допомогою моделі Midjourney, з цієї таблиці можна зробити висновок, що велику роль відіграє спосіб відображення (кольорова палітра, стиль рисунку, тощо) бажаного об'єкта.

Як можна побачити, синтезовані зображення хоча і утворилися, спираючись на іншу текстову підказку, проте до них належать всі вищенаведені невідповідності критеріям ілюстрації в інклюзивній літературі, як і для попередніх синтезованих зображень.

Висновки. Отже, на прикладі моделей Midjourney, Stable Diffusion та DALL·E 2 був проведений експеримент, який визначив, що моделі із штучним інтелектом, що перетворюють текст у зображення, не можуть використовуватися для автоматичного створення ілюстрації для інклюзивної літератури, бо такі зображення не відповідають критеріям якості інклюзивного зображення повною мірою. Проте згенеровані зображення можуть бути взяті за основу під час підготовки ілюстрації для інклюзивної літератури ілюстратором із подальшим компетентним доопрацюванням, доводячи ілюстрацію до такої, яка б відповідала критеріям якості.

Зазначимо, що синтез тексту в зображення дає змогу користувачу створювати цифрові зображення творів мистецтва за секунди, порівняно з десятками чи навіть сотнями годин традиційними методами. Ця особливість здатна вирішити одні з нагальних проблем галузі інклюзивної літератури — дефіцит компетентних кадрів із освітою в сфері образотворчого мистецтва, що мають навички підготовки випукло-тактильних ілюстрацій для інклюзивної літератури, та великі часові витрати на створення таких ілюстрацій, що пов'язано із нетривіальною технологією їх створення.

Підсумовуючи, можна зазначити, що метод синтезу ілюстрацій за допомогою моделей штучного інтелекту перетворення тексту в зображення має потенціал у вирішенні нагальних проблем сфери ілюстрації в інклюзивній літературі, проте такий метод потребує доопрацювання та подальшого дослідження.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Джуринський Є. А., Маїк В. З. *Аналіз процесу підготовки ілюстрацій для інклюзивної літератури*. Квалілогія книги. 2022. № 1 (41). С. 7−15.
2. Midjourney AI model tool for text-to-image conversion. URL: https://www.midjourney.com/ (access date: 04/05/2023).
3. Stable Diffusion AI model tool for text-to-image conversion. URL: https://stablediffusionweb.com/ (access date: 04/05/2023).
4. DALL·E 2 AI system that can create realistic images and art from a description in natural language. URL: https://openai.com/product/dall-e-2/ (access date: 04/05/2023).
5. Rombach R., Blattmann A., Lorenz D., Esser P., Ommer B. High-Resolution Image Synthesis with Latent Diffusion Models. Ludwig Maximilian University of Munich & IWR. 2022. doi: https://doi.org/10.48550/arXiv.2112.10752.
6. Ramesh A., Dhariwal P., Nichol A., Chu C., Chen M. Hierarchical Text-Conditional Image Generation with CLIP Latents. 2022. doi: https://doi.org/10.48550/arXiv.2204.06125.

7. Oppenlaender J. The Creativity of Text-to-Image Generation. In 25th International Academic Mindtrek conference (Academic Mindtrek 2022), November 16–18, 2022, Tampere, Finland. ACM, New York, NY, USA, 11 pages. 2022. doi: https://doi.org/10.1145/3569219.3569352.

REFERENCES

1. Dzhurynskyi, Ye. A., & Maik, V. Z. (2022). Analiz protsesu pidhotovky iliustratsii dlia inkliuzyvnoi literatury: Kvalilohiia knyhy, 1 (41), 7−15 (in Ukrainian).
2. Midjourney AI model tool for text-to-image conversion. Retrieved from https://www.midjourney.com/ (access date: 04/05/2023) (in English).
3. Stable Diffusion AI model tool for text-to-image conversion. Retrieved from https://stablediffusionweb.com/ (access date: 04/05/2023) (in English).
4. DALL·E 2 AI system that can create realistic images and art from a description in natural language. URL: https://openai.com/product/dall-e-2/ (access date: 04/05/2023) (in English).
5. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-Resolution Image Synthesis with Latent Diffusion Models. Ludwig Maximilian University of Munich & IWR. doi: https://doi.org/10.48550/arXiv.2112.10752 (in English).
6. Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., & Chen, M. (2022). Hierarchical Text-Conditional Image Generation with CLIP Latents. doi: https://doi.org/10.48550/arXiv.2204.06125 (in English).
7. Oppenlaender, J. (2022). The Creativity of Text-to-Image Generation. In 25th International Academic Mindtrek conference (Academic Mindtrek 2022), November 16–18, 2022, Tampere, Finland. ACM, New York, NY, USA. doi: https://doi.org/10.1145/3569219.3569352 (in English).

doi: 10.32403/1998-6912-2023-1-66-155-163

PREPARATION OF ILLUSTRATIONS FOR INCLUSIVE LITERATURE WITH THE HELP OF ARTIFICIAL INTELLIGENCE MODELS OF IMAGE FROM TEXT SYNTHESIS

Ye. A. Dzhurynskyi, V. Z. Mayik

Ukrainian Academy of Printing,
19, Pid Holoskom St., Lviv, 79020, Ukraine
vol_maik@meta.ua

A significant problem in printed inclusive literature is the preparation of illustrations, which should convey the information to the reader that the author lays down in a graphic way. Considering the fact that the target audience of inclusive literature is people with visual impairments, it is worth remembering to comply with the requirements for such illustrations, taking into account both the technical limitations of print media and the peculiarities of the human tactile and nervous system. One of the pressing problems in the field of convex-tactile illustrations is the need to spend a large amount of time

on the preparation of even one illustration. This problem is primarily related to the shortage of competent personnel with an education in the field of fine arts, who have the skills to prepare an image as a convex-tactile illustration in inclusive literature. In addition, the preparation of such an image is more likely to occur in an intuitive way, often determined by a specific editor or printer, which hinders the coherent development of illustration in inclusive literature. The tools of artificial intelligence capable of solving the given problems are considered. Among such tools there are the studies of image from text synthesis models: Midjourney, Stable Diffusion, DALL·E 2. These models are powerful tools capable of solving a wide range of problems, providing a high level of uniqueness and variability of results. The main advantage of such tools is the automation of the illustration preparation process. First of all, the automation of this process can solve one of the urgent problems of this field — the shortage of competent personnel. In addition, a big advantage of the automated approach is the significant saving of time – an illustration that could previously take hours or days to prepare in the traditional way, with the help of artificial intelligence, this task can be solved in a very short time (seconds or minutes). Experimental studies are conducted to determine the capabilities of various models of artificial intelligence for transforming text into images (illustrations) for inclusive literature, while observing the main principles and requirements for the process of preparing such illustrations. After carrying out such experiments, it is found that, despite the fact that the presented means are not able to fully prepare images as illustrations for inclusive literature, because they do not fully satisfy the requirements for such illustrations, such a method has potential and can be used in solving this problem, based on the principle of operation of such solutions.

Keywords: *visually impaired people, printing, artificial intelligence, image synthesis, text-to-image conversion, Midjourney, Stable Diffusion, DALL·E 2, information technology, inclusive technology, inclusive literature, tactile literature, image, tactile illustration, requirements for illustrations, ergonomics.*

Стаття надійшла до редакції 10.05.2023.
Received 10.05.2023.