MATHEMATICAL MODELING AND COMPUTING, Vol. 12, No. 1, pp. 283–298 (2025)

Comparison of the use of AI services based on general natural language for generating images for fiction

Yakymiv V, S.[1], Piskozub Y. Z.[1,2]

[1] *Department of Applied Mathematics, Lviv Polytechnic National University, 12 S. Bandera Str., 79013, Lviv, Ukraine*
[2] *Department of Applied Mathematics, Cracow University of Technology, Warszawska Str., 24, 31-155 Krakow, Poland*

(Received 3 September 2024; Revised 5 March 2025; Accepted 6 March 2025)

The article describes a method of image generation with Artificial Intelligence services Dall-e, MidJourney and Stable Diffusion using text abstraction retrieved with Artificial Intelligence services ChatGPT, Claude, Copilot, PI, Gemini, that work with natural language. The implementation of the new approach gives a significant gain in image quality and consistency with analysed text. The methodology is based on using neural network API services instead of commonly used natural language algorithms to extract keywords or sentences. Proposed evaluation is applied to the generated images. An analysis of evaluation options is carried out depending on Artificial Intelligence service, based on the tested book, length of result abstract, number of errors for each type and number of times AI service can understand tested book title out of proposed abstraction.

Keywords: *artificial intelligence; computing; AI-generated images; text-to-image generation.*

2010 MSC: 68T50 **DOI:** 10.23939/mmc2025.01.283

1. Introduction

AI-based services and devices are increasingly used in everyday life in various fields. Especially noteworthy are services based on neural network models for generating images based on natural language. Such systems require any text as an input parameter, that the system will convert into a graphic image. The input text can be any number of words describing the desired result. Models of these neural networks are trained on image-text pairs. As a result, selected input text is very important for the model to understand the task at hand and generate a correct result that should convey the meaning of the text as accurately as possible [1–3].

The input queries for these systems are combinations of words that are parameters, parts of objects that the user wants to see in the image [1–4]. However, will the AI model be able to cope with the task correctly if the input parameter is a large text, for example, a page from a fiction book [5]? In this case, it will be more effective to use keywords or a brief summary of the page to highlight the main objects of the scene to be displayed in the generated image.

This article is a continuation of a previous study [19] comparing the results of using natural language-based algorithms for keyword or summary extraction and AI services (currently under review). The previous study showed that images generated using abstraction obtained from AI services (ChatGPT versions of gpt-3.5-turbo and gpt-4-1106-preview models) better match the given text. Therefore, this article will compare only the AI services for obtaining abstractions that are available for use at the time of writing. The study includes comparing a sample of general text description using prepared queries to AI services and comparing the generated illustrations based on the data obtained. Based on the results, it will be possible to conclude which AI services can generate abstractions for a work of art that are better suited for generating images.

2. Related work

2.1. Text-to-Image generation

The first studies on image generation using neural networks appeared in 2015–2016 [3,6,7], although a model for generating images from a key phrase was presented in 2007 [8]. This paper describes a system for generating images based on natural language text from children's books and scientific articles. The proposed approach differs from previous models [8] as it uses AI natural language processing to generate abstraction instead of using general algorithms for keyword extraction or text summation [9, 10].

At the same time, first studies on image generation had poor quality results that often were very blurred and had primitive objects in it [3, 6, 7]. AI models available today can already produce photo-realistic quality images that are often difficult to distinguish from the real photos [11, 12].

2.2. Prompt analyzing of Text-to-Image generation systems

A lot of research has been done on analyzing input data for an AI model and output images. For instance, which input text or object description can provide an image that better matches the input query and can also provide better quality [2, 13–17]. There have been studies aimed at studying the reverse process: extracting parameters or object descriptions from AI-generated images to study the relationship between input parameters and output text in order to analyze which keywords or objects descriptions can guarantee a better result [4, 18].

3. Methodology

This section describes methodology of analyzing the comparison and performance of AI systems for generating images from abstractions using text from pages of classic literature obtained with the help of AI services. For this purpose, a list of 20 books was prepared by different authors, centuries and genres:

— The Count of Monte Cristo by Alexandre Dumas and Auguste Maquet;
— The Sign of the Four by Arthur Conan Doyle;
— Dracula by Bram Stoker;
— A Christmas Carol in Prose; Being a Ghost Story of Christmas by Charles Dickens;
— The Life and Adventures of Robinson Crusoe by Daniel Defoe;
— The Great Gatsby by F. Scott Fitzgerald;
— Metamorphosis by Franz Kafka;
— The Time Machine by H. G. Wells;
— At the mountains of madness by H. P. Lovecraft;
— Moby Dick; Or, The Whale by Herman Melville;
— The Call of the Wild by Jack London;
— Grimms' Fairy Tales by Jacob Grimm and Wilhelm Grimm;
— The Last of the Mohicans by James Fenimore Cooper;
— Pride and Prejudice by Jane Austen;
— Gulliver's Travels by Jonathan Swift;
— A Journey to the Centre of the Earth by Jules Verne;
— The Wonderful Wizard of Oz by L. Frank Baum;
— Alice's Adventures in Wonderland by Lewis Carroll;
— The Picture of Dorian Gray by Oscar Wilde;
— Treasure Island by Robert Louis Stevenson.

Using general random algorithm 5 pages were chosen from each book, receiving at the end 100 test pages.

3.1. AI text abstraction extraction

For the analysis, we used the open API of the AI systems and online chat of AI systems that are available on the market at the time of article writing.

AI services with open API:
— ChatGPT by OpenAPI with model "gpt-4o-2024-05-13";
— Claude3;

AI services by using online Web chat:
— Copilot;
— PI;
— Gemini;

In addition, there are also AI services from Meta and X. But at the time of article writing they had closed access and available only in some countries or for testers.

The API or online chat receives commands that should be executed on the proposed text. In previous research the command "You will be provided with a block of text, and your task is to return a short one sentence of what is happening in provided text." was used to generate abstraction. Previous research showed that result that's not very short generates much better images in AI services. As a result, for current study, the command was changed: "You will be provided with a block of text, and your task is to return summarization (not big) that describes provided text." With changed command results were expected to have much better relation to the text as provided command is now much clearer.

3.2. Image generation

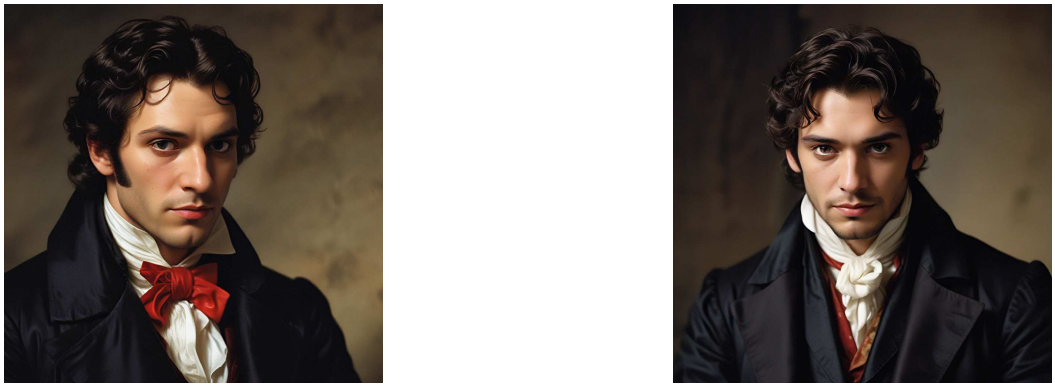
The next stage of the research is to generate images using AI services from generated short descriptions. At this stage, three AI systems that have an open API would be used and compared:
— Dall-e;
— Stable Diffusion;
— Midjourney.

The result of using the API is a URL link to the image, stored along with the keywords and the referred text description. Each API has its own configuration for the service to specify parameters for generated images. For instance, Dall-e has next configuration:
— model — "dall-e-3";
— prompt — generated abstraction from previous step;
— size — "1024 × 1024";
— quality — "standard".

The only configuration for Midjourney was the number of generated images (1), being 4 by default. The rest of configuration used in test were default parameters:
— version — "6.1";
— size — "1024 × 1024";
— mode — "standard".

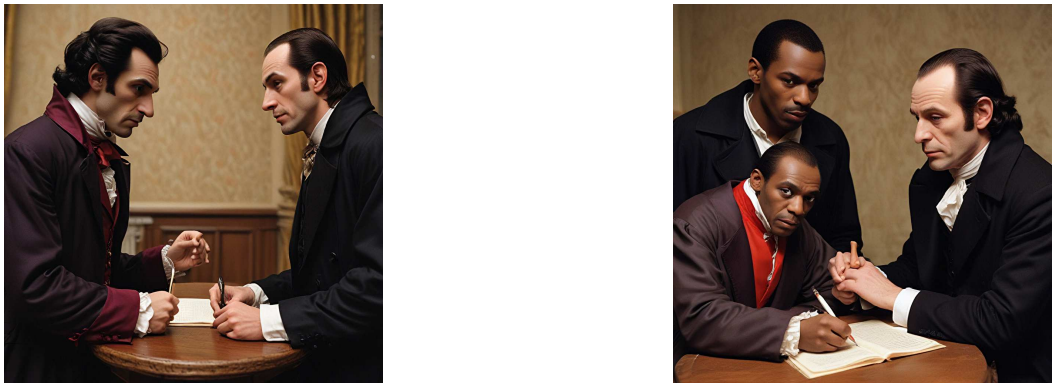
Stable Diffusion has much more configuration for API request. But the main parameters are:
— version — "3";
— negative prompt — is empty;
— guidance scale — "7.5";
— safety checker — "yes";
— size — "1024 × 1024;
— mode — "standard".

At the time of article writing, a newer version of Stable Diffusion model is available — v6. It was tested to see if there was any difference or improvements in the performance or image quality generation of the new model. But the results obtained, namely the generated images, were practically the same as those obtained using version 3 (see Figures 1–3). However, the performance and operation of the model in general were poor. Only 40% of the queries received a generated image. At the same time, the image generation time was dozen times longer. Therefore, it was decided to use version 3, similar to the previous study [19].



a ***b***

Fig. 1. Image generated by Stable Diffusion v3 (***a***) vs v6 (***b***) using Copilot text abstraction: "In this excerpt from the novel 'The Count of Monte Cristo,' the author describes a confrontation between the Count and Caderousse. The Count, who was once Caderousse's accomplice, demands that he write a letter revealing the truth about a man intending to marry the banker Danglars' daughter. Despite Caderousse's fear, he complies with the Count's request. This passage highlights tension, intrigue, and the Count's mysterious nature."



a ***b***

Fig. 2. Image generated by Stable Diffusion v3 (***a***) vs v6 (***b***) using Gemini text abstraction: "The Count overpowers a thug who tries to stab him, forces him to write a confession to ruin another man, and then lets him go."

a ***b***

Fig. 3. Image generated by Stable Diffusion v3 (***a***) vs v6 (***b***) using GPT-4 text abstraction: "Mercedes, engaged in a poignant conversation with Edmond (Monte Cristo), expresses that she desires nothing for herself as she lives between the memories of her lost love, Edmond Dantes, and her deceased husband, whom Edmond killed. She reveals her intention to live a life of prayer sustained by a small sum of money she found, shunning the larger fortune left by her late husband. Despite Edmond's offers to provide for her and ensure her son's approval, Mercedes remains resigned, expressing her passive submission to fate and her hope of reunion in heaven. Ultimately, she remains steadfast in her solitary grief as Monte Cristo departs."

3.3. Image and abstraction scoring

The following scale was proposed to evaluate the received text description (see Table 1).

Table 1. Scale for assessing the relevance of keywords or descriptions to the text.

Text mark	Number mark	Description
Not relevant	1–2	The provided keywords or summary have minimal to no connection with the content of the text. There's little or no overlap in terms of theme or information.
Slightly Relevant	3–4	There are some minor connections between the keywords or summary and the text, but the correspondence is weak. The keywords may touch on peripheral aspects.
Somewhat Relevant	5–6	There is a moderate level of relevance. The keywords or summary capture some aspects of the text, but there are notable gaps or differences.
Moderately Relevant	7–8	The keywords or summary align well with the text, capturing the main ideas and themes. However, there may be some nuances or details that are not perfectly reflected.
Highly Relevant	9	The keywords or summary closely match the content of the text. They effectively encapsulate the main points and themes, with only minor variations.
Perfectly Relevant	10	The keywords or summary perfectly describe the text. Every important detail, theme, and nuance is accurately reflected.

The same scale was used to assess the relevance of images to text.

4. Results and discussions

The key element of this study, along with the evaluation and analysis of the results, is that it examines not only how accurate the keywords are for the text analyzed, but also how relevant they are for further use as a query for image generation.

The evaluation was conducted only by the author of the research, so some subjective discrepancy in the obtained scores should be taken into account.

4.1. Text abstraction score results analysis

Based on the evaluation data obtained for each individual page of each book, the average value was calculated for each abstraction obtained, as well as for the images generated by the AI. The average evaluation values for textual descriptions for the books under study are shown in table (see Table 2).

Table 2. Average score for book by AI service.

A. Dumas and A. Maquet — The Count of Monte Cristo	9.2	8.6	7.8	9.6	8.6
A. C. Doyle — The Sign of the Four	8.4	8	8.2	8.8	8.8
B. Stoker — Dracula	8.6	7.8	8.2	9.8	8.4
C. Dickens — A Christmas Carol in Prose; Being a Ghost Story of Christmas	8.4	8.4	9	9.8	8.6
D. Defoe — The Life and Adventures of Robinson Crusoe	9	8.8	9.4	9.4	9
F. S. Fitzgerald — The Great Gatsby	8.8	7.8	9.2	9.8	8.4
F. Kafka — Metamorphosis	8.2	8	8.4	8.4	8.8
H. G. Wells — The Time Machine	8.8	8	8.6	9.2	8.4
H. P. Lovecraft — At the mountains of madness	8.8	8	8	9.6	9
H. Melville — Moby Dick; Or. The Whale	9.2	8.2	9	9.6	8.8
J. London — The Call of the Wild	8.8	8.4	8.4	9.2	9
J. Grimm and W. Grimm — Grimms' Fairy Tales	8.8	8.8	8.8	9.4	7.8
J. F. Cooper — The Last of the Mohicans	8.8	8	9.4	9.2	8.8
J. Austen — Pride and Prejudice	9.6	9	8.8	9.2	9.8
J. Swift — Gullivers Travels	8.6	7	8	8.6	6.6
J. Verne — A Journey to the Centre of the Earth	8.8	7.6	8.6	9	6.4
L. F. Baum — The Wonderful Wizard of Oz	9.2	9	9	9.4	8.8
L. Carroll — Alice's Adventures in Wonderland	8.8	8.2	9.4	9.4	9
O. Wilde — The Picture of Dorian Gray	7.2	9	7.4	8.8	8.6
R. L. Stevenson — Treasure Island	8	7.8	8	8.2	8

Based on the results presented in table (see Table 2) and the diagram (see Figure 4), it can be seen that the results of all methods are quite close with a slight difference. However, in most cases, the results of GPT-4 are ahead of the other approaches. It is difficult to single out the lowest performing methods because the bottom scores differ from one book to another. Nevertheless, about half of the lowest scores were achieved by Copilot. The AI service Pi also received some of the lowest scores. Claude3 and Gemini showed average results, slightly lower than GPT-4.

However, this study aims to analyze which approach would be better as an input text for an AI image generation service. Therefore, these results are rather superficial, as it is possible that the AI service will be able to determine on its own whether the keywords belong to a particular novel or to a place or character and generate the corresponding image.

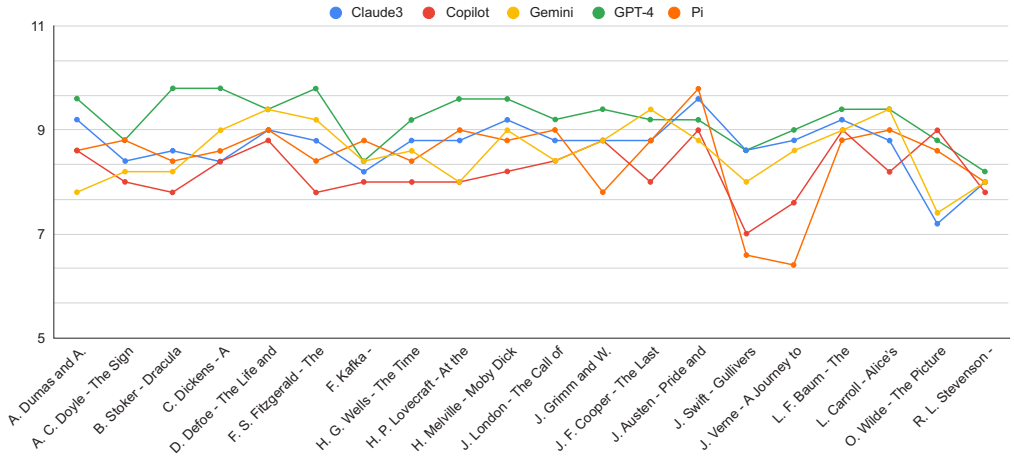


Fig. 4. Average abstraction score for book by AI service.

4.2. Book and author recognition

The generated abstracts often contained the title of the original work or text characters, despite the fact that the tested text did not contain such details as the author, title of the work or when main characters were not mentioned in the text. This means that AI service was trained on specific books that were also used in current research, and as a result, it could recognize the book in a few paragraphs (Figure 5).

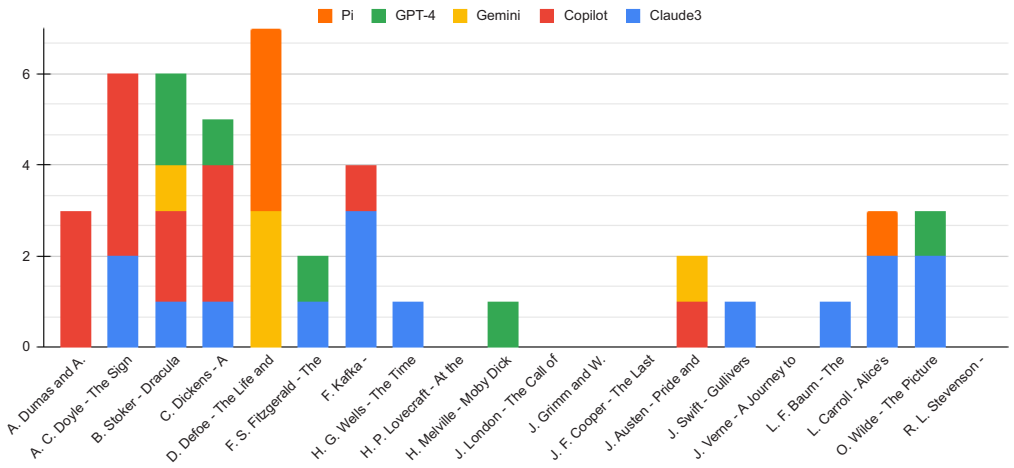


Fig. 5. Number of text recognitions by AI service.

In total, Claude3 recognized 15 pages from tested books, namely about 33% of the total number of recognitions; Copilot recognized 14 pages from different books, namely about 14%; GPT-4 recognized 6 pages, namely 13.3%; Pi and Gemini recognized 5 pages, namely 11.1% of the total number of recognitions. It is worth noting that 70% of the tested books were recognized, that means some AI services have been trained on many specific books, including literature, but not all of them.

Examples of abstractions obtained when a book title is mentioned, generated by Copilot and Claude3 respectively:

> In this dramatic scene from "Dracula" by Bram Stoker, Van Helsing and his allies confront a transformed Lucy Westenra in a cemetery. Having prepared with the Host from Amsterdam, they witness Lucy, now a vampire, clutching a child at her breast. The moonlight reveals her change from pure to cruel, with her lips stained with blood. Despite their shared shock and Arthur's near-collapse, they resolutely face the monstrous figure that was once Lucy, embodying their transition from disbelief to grim acceptance of their dire situation.

> This passage describes the climactic moment and resolution of Charles Dickens' "A Christmas Carol." Ebenezer Scrooge, confronted with his own mortality and the consequences of his actions, experiences a profound transformation, vowing to change his ways and embrace the spirit of Christmas in his heart all year round.

These results demonstrate that AI can "guess" the tested book database from the text if it has been trained on it. Generated images with abstraction that has book title mentioned were additionally compared to images generated using the same abstractions but with book title mentioning removed. It is hard to say whether the presence of a book title (and thus a more detailed description) really affects the quality of the image. In the examples of images presented (see Figure 6–7), it can be seen that in some cases the absence of a title resulted in images that corresponded to the description somewhat better (more detailed scene, more accurate number of characters depicted). However, in other cases (see Figure 8–9), the images contain unnecessary details or are less relevant to the description.

a ***b*** ***c*** ***d***

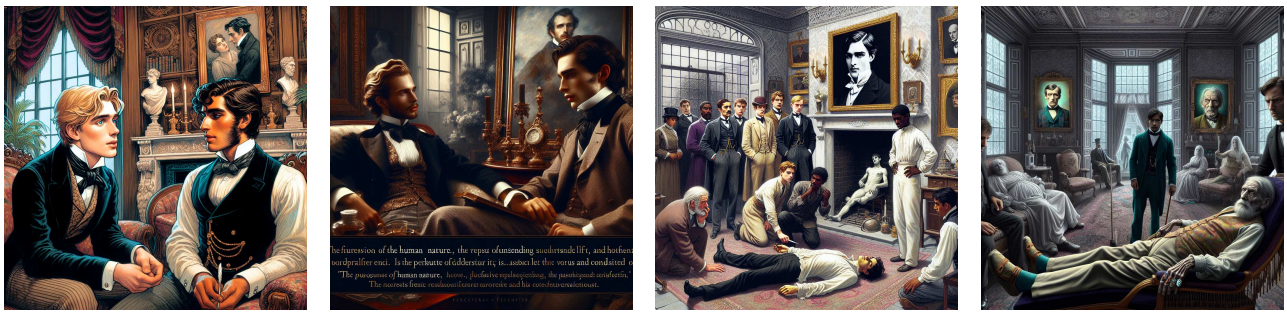
Fig. 6. Generated images by Dall-e for Alexandre Dumas "The Count of Monte Cristo" using Copilot abstraction (***a***), (***c***) with book title mentioned in abstraction (***b***), (***d***) without book title.

a ***b*** ***c*** ***d***

Fig. 7. Generated images by Dall-e for Arthur Conan Doyle "The Sign of the Four" using Claude3 abstraction (***a***) with book title mentioned in abstraction (***b***) without book title and using Copilot abstraction (***c***) with book title mentioned in abstraction (***d***) without book title.

a ***b*** ***c*** ***d***

Fig. 8. Generated images by Dall-e for Franz Kafka "Metamorphosis" using Claude3 abstraction (***a***), (***c***) with book title mentioned in abstraction (***b***), (***d***) without book title.



a ***b*** ***c*** ***d***

Fig. 9. Generated images by Dall-e for Oscar Wilde "The Picture of Dorian Gray" using Claude3 abstraction (***a***) with book title mentioned in abstraction (***b***) without book title and using GPT-4 abstraction (***c***) with book title mentioned in abstraction (***d***) without book title.

4.3. Image generation errors analysis

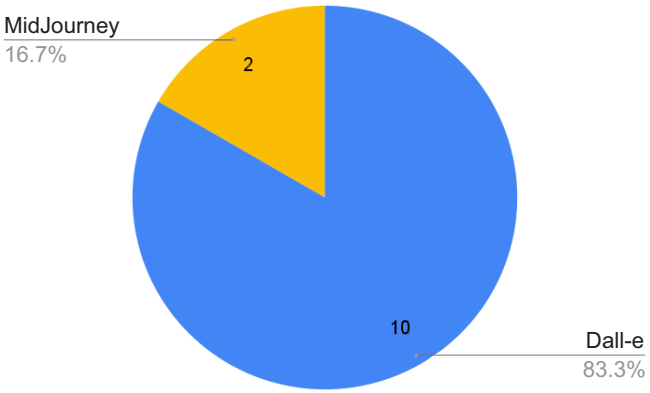


Fig. 10. Number of errors from AI image generation service.

Table 3. Number of errors for tested AI services.

	Dall-e	Stable Diffusion	MidJourney
Claude3	2	0	0
Copilot	2	0	1
Gemini	1	0	1
GPT-4	3	0	0
Pi	2	0	0
Summary	10	0	2

It is worth noting that the generation of abstractions of the analyzed page by AI services contains problems that would not arise if other methods were used, such as the use of natural language algorithms.

The API is designed to repeat the request in case of an error in order to minimise the risk of possible Internet outages, internal system errors, or false alarms (in the case of Dall-e's input word filters).

In short, AI services have a list of unacceptable words that are used to check queries sent by users. The list of such words include racist, politically incorrect, or swear words.

Most of these errors were received when using the DALL-e service (Figure 10). DALL-e also refused to process requests that contained descriptions of bloody scenes. For example, most of these cases occurred when generating images for randomly selected pages from H. P. Lovecraft's "In the Mountains of Madness." Other image generation services did not indicate the cause of the errors.

Comparing AI models, the largest number of errors was obtained using abstraction generated by GPT-4 model. The summary statistics of errors obtained during image generation by Dall-e, Stable Diffusion, and MidJourney services are presented in table (see Table 3).

4.4. Image generation score results analysis

This section presents and analyses examples of generated images (see Figures 11–15) for each of the services used (Dall-e (***a***), MidJourney (***b***), Stable Diffusion (***c***)), which, according to the evaluation results, best match the selected text. That is the randomly selected page from the novel "Pride and Prejudice", by Jane Austin. Tested page describes two women Elizabeth and Mrs. Gardiner having a conversation about Lydia and Mr. Wickham.

For instance, the image generated by Dall-e (Figure 11*a*) using Claude3 abstraction looks good and related to context on the text: the image contains all described characters mentioned in the text and in general looks like characters are having a conversation, mentioning other people. The image created with MidJourney (Figure 11*b*) also have good relevance to the text and includes references to some of the main characters. StableDiffusion image (Figure 11*c*) have no relevance at all without any connection to the content of the text.

Abstraction generated by Copilot is a bit shorter. Nonetheless, all images contain main characters having a conversation (see Figure 12). But there are more people in the image generated by Dall-e (see Figure 12*a*) than in the text with additional details. MidJourney and StableDiffusion images have good relevance to the text, including references to the main characters.

a ***b*** ***c***

Fig. 11. Generated images for Claude3 generated abstraction "This passage describes Elizabeth's concerns about her sister Lydia's behavior and character, as well as her negative opinion of Mr. Wickham. Elizabeth expresses worry about Lydia's frivolous nature and susceptibility to Wickham's charms, while also revealing her own knowledge of Wickham's true, deceitful character to her aunt, Mrs. Gardiner."

Abstraction generated by Gemini is also relatively short. As a result, generated images (see Figure 13) looks similar to Copilot images, where Dall-e image has some extra details.

a ***b*** ***c***

Fig. 12. Generated images for Copilot generated abstraction "Elizabeth discusses Lydia's behavior and Wickham's character with Mrs. Gardiner. She expresses concern about Lydia's lack of seriousness and Wickham's deceitful nature. The story highlights the consequences of Lydia's actions and the family's reactions."

a *b* *c*

Fig. 13. Generated images for Gemini generated abstraction "Elizabeth defends Jane's innocence regarding Wickham. Elizabeth reveals she knows Wickham is a liar and has proof of his deception towards Mr. Darcy. She regrets not telling Lydia the truth about Wickham sooner."

a *b* *c*

Fig. 14. Generated images for GPT-4 generated abstraction "Elizabeth, deeply distressed, discusses her younger sister Lydia's lack of decency and virtue due to her upbringing focused on amusement and vanity. She is concerned about Lydia's infatuation with officers, particularly the deceitful Wickham, who Elizabeth and Jane know to be dishonorable and profligate. Elizabeth regrets not revealing Wickham's true character earlier, fearing that now Lydia may be in danger due to her ignorance of his deceit."

Abstraction generated by GPT-4 is much longer comparing to previous and is mentioning only one person from the text. Therefore, images generated by MidJourney and Stable Diffusion (see Figures 14*b*, 14*c*) contain only one mentioned character, while Dall-e image has 2 characters but again with extra details in the scene.

a *b* *c*

Fig. 15. Generated images for Pi generated abstraction "Elizabeth discussed Lydia's elopement with Mrs. Gardiner, lamenting that Lydia, who had been raised with no concept of morality, had been easily seduced by Wickham's charm. Elizabeth had previously heard about Wickham's dishonorable behavior, and knew he was not to be trusted, but had failed to warn Lydia. Mrs. Gardiner wondered how Lydia could be so oblivious to Wickham's true character, but Elizabeth could only explain that Lydia had been sheltered from the truth and, thus, was easily fooled by Wickham."

Finally, abstraction generated by Pi AI model is also big. All images created by AI models (see Figure 15) show a strong connection with the text containing the conversation of the main characters. Dall-e image still have additional details in the form of text in the picture.

Table 4 shows the average results for all tested image evaluation pages generated by Dall-e for each approach and book. Figure 16 shows that it is difficult to isolate the approach that provided the best scores for the most relevant image.

Table 4. Average image score for book by Dall-e.

	Claude3	Copilot	Gemini	GPT-4	Pi
A. Dumas and A. Maquet — The Count of Monte Cristo	7.4	6.6	6	7.6	6.8
A. C. Doyle — The Sign of the Four	7	7.8	7.8	7.4	7.8
B. Stoker — Dracula	6.75	6.8	7.2	7.2	6.75
C. Dickens — A Christmas Carol in Prose	7.6	7.6	7.6	7.8	7.6
D. Defoe — The Life and Adventures of Robinson Crusoe	7.2	6.2	7.4	6.8	7.8
F. S. Fitzgerald — The Great Gatsby	7	7.4	7.2	6.6	6.6
F. Kafka — Metamorphosis	6.4	6.6	6	6.4	6.6
H. G. Wells — The Time Machine	7.2	7.2	7.4	7.4	7.4
H. P. Lovecraft — At the mountains of madness	6.25	6.25	6.5	6.75	6
H. Melville — Moby Dick; Or, The Whale	7.2	7.2	6.8	7.6	7.2
J. London — The Call of the Wild	7.4	7	7.2	6.8	7.25
J. Grimm and W. Grimm — Grimms' Fairy Tales	7.4	7.25	7.6	6.5	6.2
J. F. Cooper — The Last of the Mohicans	6.8	6	7.4	6.4	6.6
J. Austen — Pride and Prejudice	8	8.4	7.8	8.4	8
J. Swift — Gullivers Travels	5.2	5.6	5.6	6	5.4
J. Verne — A Journey to the Centre of the Earth	6.8	6	5.8	6.4	6.4
L. F. Baum — The Wonderful Wizard of Oz	6.4	7.2	6.2	6.2	6.2
L. Carroll — Alice's Adventures in Wonderland	8	7.2	7.8	6.75	7.2
O. Wilde — The Picture of Dorian Gray	5.8	7.6	6.2	7	7.4
R. L. Stevenson - Treasure Island	6.8	5.2	6.8	6.8	7.2

Only one of the results has a score of more than 8. Although the chart shows that more than half of the highest scores are for images generated by the GPT-4 model. In all other cases, other models have better results depending on the book. The lowest scores have images generated by all models for "Gullivers Travels" by J. Swift scores.

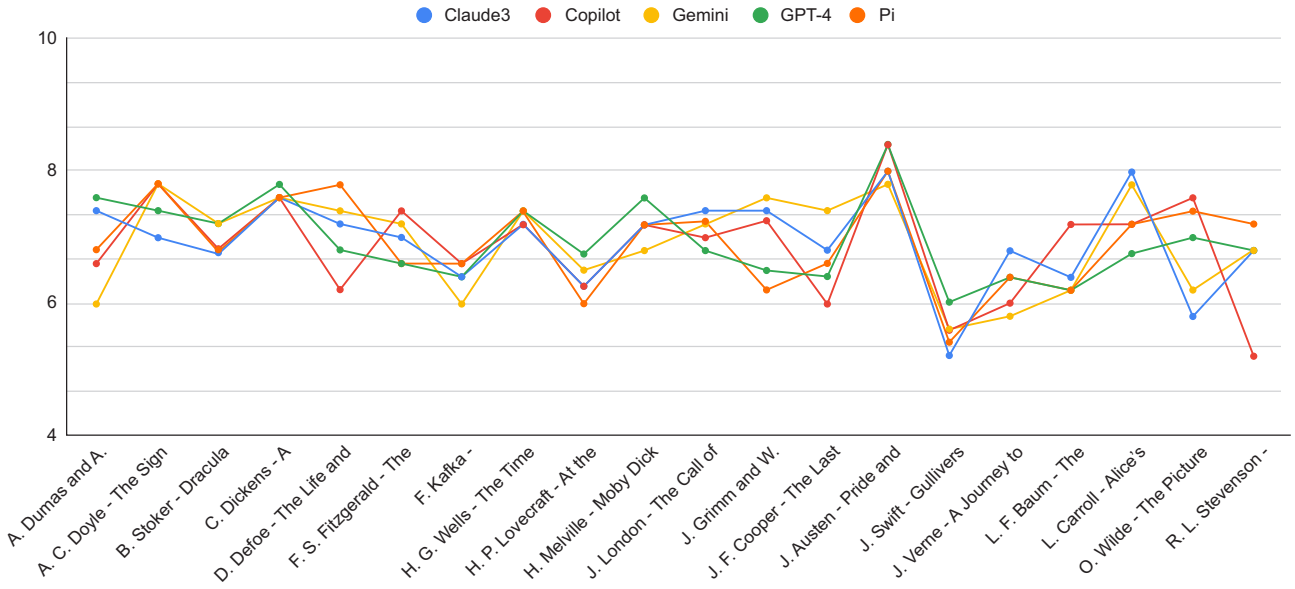


Fig. 16. Average image score for book by Dall-e.

Table 5. Average image score for book by MidJourney.

	Claude3	Copilot	Gemini	GPT-4	Pi
A. Dumas and A. Maquet — The Count of Monte Cristo	7.2	7.2	6.75	7.6	7.2
A. C. Doyle — The Sign of the Four	7.2	7.8	7.4	7.8	7.6
B. Stoker — Dracula	6.6	7.4	7.6	7.4	6.8
C. Dickens — A Christmas Carol in Prose	8	7.6	8	8.2	7.8
D. Defoe — The Life and Adventures of Robinson Crusoe	8.2	8	8.4	7.8	8.4
F. S. Fitzgerald — The Great Gatsby	7.4	7.2	7.6	7.2	6.8
F. Kafka — Metamorphosis	7	6.8	6.4	6.6	6.6
H. G. Wells — The Time Machine	8.2	7.6	7.6	7.8	7.2
H. P. Lovecraft — At the mountains of madness	6.8	6.4	6.6	6.4	6.6
H. Melville — Moby Dick; Or, The Whale	6.6	6.6	6.4	6.6	6.4
J. London — The Call of the Wild	7.6	7.8	6.8	7.2	6.8
J. Grimm and W. Grimm — Grimms' Fairy Tales	7.2	7.6	7.4	6.8	7
J. F. Cooper — The Last of the Mohicans	8.4	7.4	8.4	7	8
J. Austen — Pride and Prejudice	8.8	9	9	8.8	8.8
J. Swift — Gullivers Travels	5	5.6	6.2	5.6	5.8
J. Verne — A Journey to the Centre of the Earth	7.6	7	6.8	7.4	7.6
L. F. Baum — The Wonderful Wizard of Oz	8.4	8	7.8	8	7
L. Carroll — Alice's Adventures in Wonderland	7.8	8.2	7.6	7.6	7.2
O. Wilde — The Picture of Dorian Gray	6.4	7.8	6.6	8.8	8.6
R. L. Stevenson - Treasure Island	7.8	8	9	7.6	7.2

The results of MidJourney image evaluation are presented in Table 5. The diagram (see Figure 17) shows similar results to Dall-e, with large jump for "Pride and Prejudice" by J. Austen scores and a drop for "Gullivers Travels" by J. Swift scores. At the same time, MidJourney's images has more scores above 8, indicating that MidJourney has better overall performance than Dall-e.

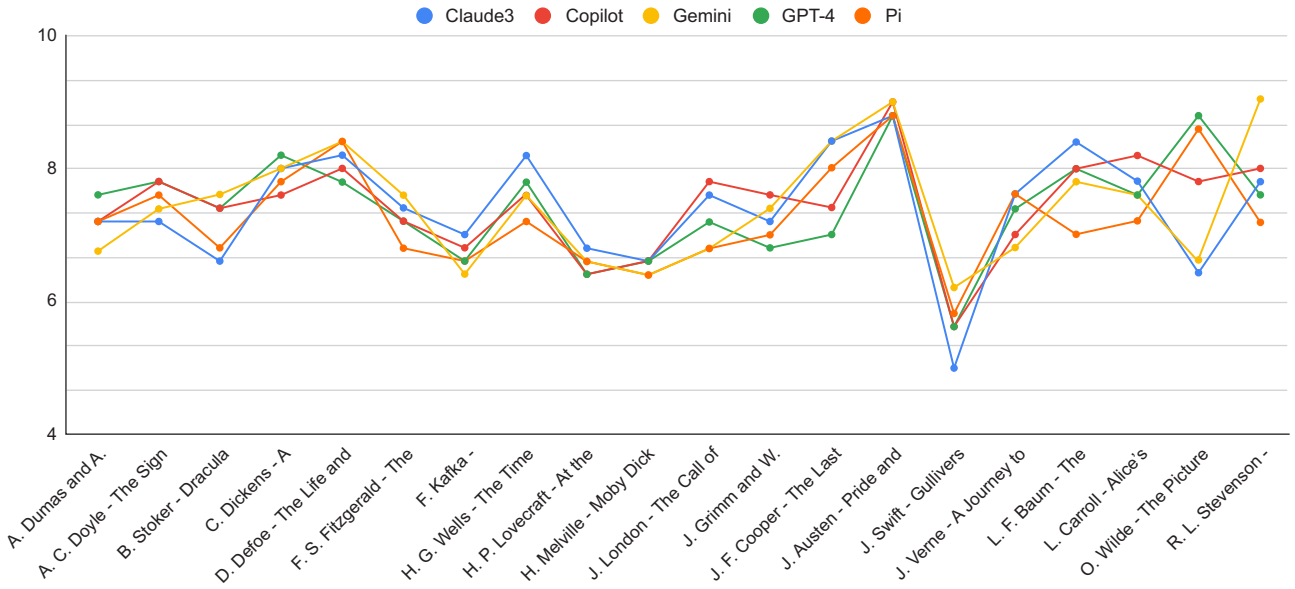


Fig. 17. Average image score for book by MidJourney.

Table 6 contains the average evaluation results for images generated by the Stable Diffusion service and the diagram (see Figure 18) represents the data from the table. The average scores on the diagram looks worse compared to previous MidJourney and Dall-e scoring, but with similar spike for "Pride and Prejudice" by J. Austen, except for Claude3, which have also worse results compared to other AI services. The ChatGPT and Copilot have highest scores among all tested AI services.

Table 6. Average image score for book by StableDiffusion.

	Claude3	Copilot	Gemini	GPT-4	Pi
A. Dumas and A. Maquet — The Count of Monte Cristo	7.6	7.2	7.6	7.2	6.6
A. C. Doyle — The Sign of the Four	6.4	7	7	7.2	7
B. Stoker — Dracula	6	7	7	6.6	6.6
C. Dickens — A Christmas Carol in Prose	6.8	6.8	7.4	7	6.6
D. Defoe — The Life and Adventures of Robinson Crusoe	6.4	6.4	8	6.8	8.2
F. S. Fitzgerald - The Great Gatsby	4	6.8	6.8	6.8	6.6
F. Kafka — Metamorphosis	5.2	6.4	6.2	6.4	6.2
H. G. Wells — The Time Machine	7.4	7.2	7.4	7	7
H. P. Lovecraft — At the mountains of madness	6.2	6	6.2	6.2	5.8
H. Melville — Moby Dick; Or, The Whale	5.4	6.4	5.8	6.4	5.4
J. London — The Call of the Wild	5.8	7	6.4	6.6	5.4
J. Grimm and W. Grimm — Grimms' Fairy Tales	6.4	6	6	6.4	6
J. F. Cooper — The Last of the Mohicans	7.2	6	7	6.4	6.6
J. Austen — Pride and Prejudice	5	8.6	8	8.2	8.8
J. Swift — Gullivers Travels	4.8	5	5.2	5.8	4.4
J. Verne — A Journey to the Centre of the Earth	5.4	6.4	6.4	6.6	6.2
L. F. Baum — The Wonderful Wizard of Oz	6.2	6.6	6.4	6	6.2
L. Carroll — Alice's Adventures in Wonderland	7	7	7	7.2	6.2
O. Wilde — The Picture of Dorian Gray	5.2	6.8	5.6	7.8	7.2
R. L. Stevenson — Treasure Island	5	7.2	5.8	6.4	6.2

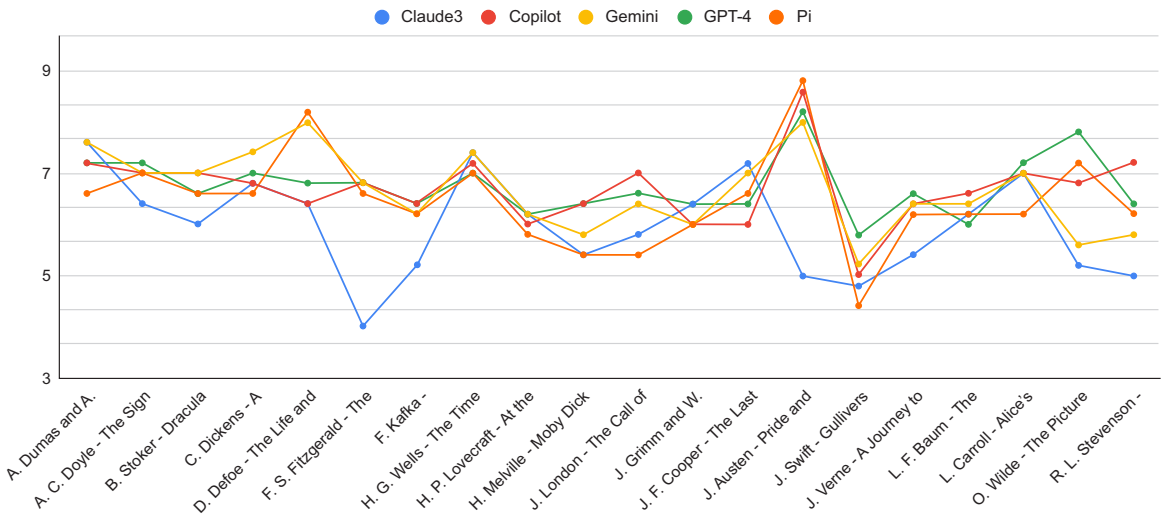


Fig. 18. Average image score for book by StableDiffusion.

Analyzing the data on the dependence of the average image score of different services on the applied algorithm, shown in the graph (see Figure 19), we can see that there is a clear difference between the results of using image generation services.

The best results were obtained for images generated by MidJourney, while Dall-e has worse results with a slight difference and StableDiffusion has lowest results. At the same time, almost all the indicators of the used

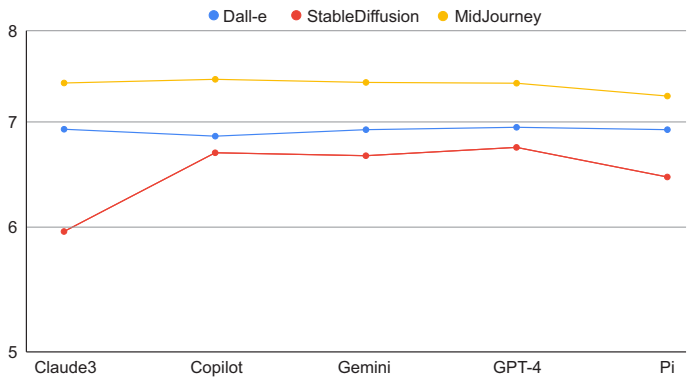


Fig. 19. Average image scores between Dall-e, MidJoruney and StableDiffusion to AI service.

AI services are at the same level for the used image generation service, except for Claude3 for Stable-Diffusion, which demonstrates the worst performance among the tested ones.

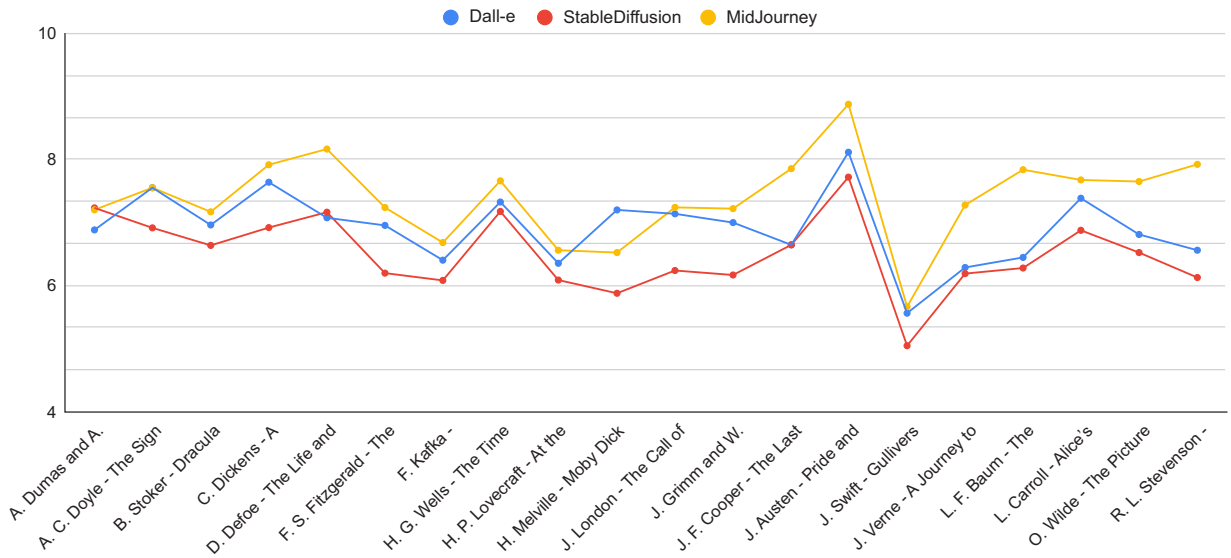


Fig. 20. Average image scores between Dall-e, MidJoruney and StableDiffusion to book.

The graph representing the dependence of the AI service evaluation results on the book (see Figure 20) MidJourney showed the highest results for almost all tested books (except for "Moby Dick; Or, The Whale" by H. Melville and "The Count of Monte Cristo" by A. Dumas and A. Maquet), while Dall-e has similar but slightly worse results, and StableDiffusion has very similar results to Dall-e, but slightly worse results.

At the beginning of the data analysis, we noticed that the generated image did not always correspond to the quality and relevance of the description, which better conveys the essence of the analyzed page.

On the contrary, the shorter the text is, the better the image is, and the fewer descriptive phrases and the more keywords are in the text, the better and more relevant the image is. Also, the generated images did not always contain all the objects that were specified in the list of keywords or in the description.

In most cases, the generated images contained only a few of the objects specified, and the rest were ignored by the AI service.

As result, we can conclude that at this stage of development of image generation systems, services have limitations in terms of full compliance of the generated images with the provided input query.

5. Conclusions

After analyzing the results, we can conclude that the described method and the commands used cope with the task, although not perfectly. The generated abstracts contain too general a description of the main essence of the text, and it is difficult for AI services to focus on the overall picture.

Comparing the results of the AI services of the generated abstractions as well as the images generated with the corresponding descriptions, we can see that ChatGPT has the best results in most cases, Claude3 and Gemini show slightly lower results, while Copilot and Pi have the lowest results. Improvements to existing language models are happening quite frequently and at a frantic pace. Also, new language models working with natural text are appearing on the market, so even in the near future, we can expect clear improvements in this area. Greater use of diverse literature for model training will, in turn, improve the results of AI language models.

In general, analyzing the data, we can notice that the images generated by StableDiffusion have the worst relevance to the text compared to the other test images. There were also frequent cases when the

generated images did not contain any objects from the used abstraction at all. When comparing the new version of StableDiffusion, there was no improvement in image matching, but only a deterioration in performance. Unless there are significant improvements in future versions of the model, continuing to study using this model will always produce the worst results.

At the same time, Dall-e and MidJourney provided a generally similar level of image quality and relevance to the text, while still allowing for guesswork by adding key details to the images when they were missing from the provided abstraction. Overall, however, MidJourney provided a better level of image quality and greater relevance to the resulting abstracts.

To improve the described approach, it is worth trying in future studies:

— use this approach for individual paragraphs to obtain more precise images that focus on fewer events and objects described;
— use AI services with a different number of keywords or limit the number of words in the generated description;
— change a given command so that the AI tries to separate scenes from a page or separate events from a page of text, and generate images for each of them;
— check and test API for the other AI services with same functionality (if there are such with open API), that will not be limited by general restrictions (i.e. political correctness).

It is worth noting, that the assessment of the appropriateness or similarity of the words chosen, the abstractions and images generated is quite subjective, because any reader can imagine and perceive the text in their own subjective way, but the task of this study is not to guess absolutely accurately and convey the formed idea of a particular page before the reader imagines it, but to help imagine, suggest or form a certain direction of images and objects to form a better perception of the work.

[1] Park D. H., Azadi S., Liu X., Darrell T., Rohrbach A. Benchmark for Compositional Text-to-Image Synthesis. NeurIPS Datasets and Benchmarks (2021).

[2] Brade S., Wang B., Sousa M., Oore S., Grossman T. Promptify: Text-to-Image Generation through Interactive Prompt Exploration with Large Language Models. UIST '23: Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology. 96, 1–14 (2023).

[3] Mansimov E., Parisotto E., Ba L. J., Salakhutdino R. Generating Images from Captions with Attention. Preprint arXiv:1511.02793 (2016).

[4] Xinyue S., Yiting Q., Michael B., Yang Z. Prompt Stealing Attacks Against Text-to-Image Generation Models. Preprint arXiv:2302.09923 (2023).

[5] Yakymiv V., Piskozub L., Piskozub, Y. Using Artificial Intelligence to Generate Real-time Augmented Reality Content in Book Publishing. International Journal of Control Systems and Robotics. **8**, 1–5 (2023).

[6] Liu S. Text-To-Image Generation (2019).

[7] Reed S., Akata Z., Logeswaran L., Schiele B., Lee H. Generative Adversarial Text to Image Synthesis. Preprint arXiv:1605.05396 (2016).

[8] Zhu X., Goldberg B. A., Eldawy M., Dyer C. R., Strock B. A text-to-picture synthesis system for augmenting communication. Proceedings of the Twenty-Second AAAI Conference on Artificial Intelligence. 1590–1595 (2007).

[9] Mihalcea R., Tarau P. TextRank: Bringing Order into Text. Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing. 404–411 (2004).

[10] Turney P. D. Learning Algorithms for Keyphrase Extraction. Information Retrieval. **2**, 303–336 (2000).

[11] Sanghyuck N., Mirae D., Kyeonah Y., Juntae K. Realistic Image Generation from Text by Using BERT-Based Embedding. Electronics. **11** (5), 764 (2022).

[12] Mathesul S., Bhutkar G., Rambhad A. AttnGAN: Realistic Text-to-Image Synthesis with Attentional Generative Adversarial Networks. Sense, Feel, Design. 397–403 (2022).

[13] Yutong X., Zhaoying P., Jinge M., Jie L., Qiaozhu M. A Prompt Log Analysis of Text-to-Image Generation System. WWW '23: Proceedings of the ACM Web Conference 2023. 3892–3902 (2023).

[14] Fan L., Wang H., Zhang K., Pei Z., Li A. Towards an Automatic Prompt Optimization Framework for AI Image Generation. HCI International 2023 Posters. 405–410 (2023).

[15] Liu V., Chilton L. Design Guidelines for Prompt Engineering Text-to-Image Generative Models. CHI '22: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. 384, 1–23 (2022).

[16] Frolov S., Hinz T., Raue F., Hees J., Dengel A. Adversarial text-to-image synthesis: A review. Neural Networks. **144**, 187–209 (2021).

[17] Hao Y., Chi Z., Dong L., Wei F. Optimizing Prompts for Text-to-Image Generation. Preprint arXiv:2212.09611 (2022).

[18] Xingqian X., Jiayi G., Zhangyang W., Gao H., Irfan E., Humphrey S. Prompt-Free Diffusion: Taking "Text" out of Text-to-Image Diffusion Models. Preprint arXiv:2305.16223 (2023).

[19] Yakymiv V., Piskozub Y. Research on the use of AI for Selecting Abstractions for Natural Language Image Generation Tools. International Journal of Computing. **23** (4), 637–654 (2024).

Порівняння використання сервісів штучного інтелекту на базі загальної природної мови для генерування зображення для художньої літератури

Якимів В. С.[1], Піскозуб Й. З.[1,2]

[1] *Кафедра прикладної математики, Національний університет "Львівська політехніка", вул. С. Бандери, 12, 79013, Львів, Україна*
[2] *Кафедра прикладної математики, Краківська політехніка, вул. Варшавська, 24, 31-155, Краків, Польща*

У статті описано метод генерації зображень за допомогою сервісів Штучного Інтелекту (ШІ) Dall-e, MidJourney та Stable Diffusion з використанням текстових абстракцій, отриманих за допомогою сервісів ШІ ChatGPT, Claude, Copilot, PI, Gemini, які працюють з природною мовою. Впровадження нового підходу дає значний виграш у якості зображення та узгодженості з аналізованим текстом. Методологія базується на використанні нейромережевих API-сервісів замість загальноприйнятих алгоритмів природної мови для вилучення ключових слів або речень. Запропонована оцінка застосовується до згенерованих зображень. Аналіз варіантів оцінювання проводиться залежно від сервісу ШІ, на основі тестованої книги, довжини анотації результату, кількості помилок для кожного типу та кількості разів, коли сервіс ШІ здатний зрозуміти тестовану книгу із запропонованої абстракції.

Ключові слова: *штучний інтелект; обчислення; AI-зображення; генерація зображень; перетворення тексту в зображення.*