

Graphical use of AI

In the June VTEI issue, we got familiar with the AI tool ChatGPT in the form of an “interview”. We continue with the topic of artificial intelligence and this time we present experiences with a more “visual” tool. Our intention was to create different visualizations of the situation using text input, the so-called “prompt”, or from a master photo, for example a watercourse restoration or the idea of building a water tower in the countryside. But before we get to the visualizations themselves, let us say a few words about this topic.

There are several AI tools that allow users to generate desired images based on text inputs, called prompts. These tools use advanced machine learning technologies and generative models and can create realistic images based on the description provided by the user. Such tools include, for example, DALL-E from the OpenAI company and MidJourney from David Holz’s American company of the same name. These tools have the potential to be used for various applications, including the creation of visual content, visual design, or even the design of new products.

For our purposes, we chose MidJourney, a service for generating graphics using artificial intelligence. The tool launched in the middle of 2022, and users create graphics using commands given to a chatbot in the Discord app.

MidJourney’s function is to recognize the relationship between images and text, where a machine learning algorithm is trained on a large number of images with text descriptions. If the user enters a request/prompt in the chat window, artificial intelligence will allow the creation of an image that matches the description.

We tested the functioning/usage of the MidJourney AI tool on four examples.

Design for the Jezerka stream restoration

In this case, the basis was the image of the visualization of the restoration of the inflow into a reservoir with a bridge and wetland vegetation published in this year’s April VTEI issue [1]. The entire process of generating the result took place in the following order – uploading a real photo of the park before the restoration (*Fig. 1a*), generating a bridge over the stream (only about 20th prompt with a satisfactory result, *Fig. 1b*), connecting both outputs (*Fig. 1c*), and fine-tuning the resulting image (*Fig. 1d*). The time required for this process was approximately three hours.

Water tower

In another case, our intention was to depict the construction of a water tower. Here, too, the source was a picture from an article on water towers published in VTEI 6/2022 (*Fig. 2a*) [2]. The following written prompt was used: “a tall concrete



Fig. 1a, b, c, d. The Jezerka stream, the situation of the groundwater drainage outlet as an occasional inflow into the water reservoir (Photo: T. Hrdinka, subsequent editing with the MidJourney tool)



Fig. 2a, b. Tower reservoir in Kolín designed by architect František Janda in the functionalist style (Photo: O. Cívín, subsequent editing with the MidJourney tool)

tower with a metal dome of the tower, featured on cg society, danube school, arial shot, watertank, germany, low pressure system, awe – inspiring award – winning, waterdrops, manufactured in the 1920s, aquiline features, parks and monuments, brenizer method --v 5” which then drew a preview image of the four variants (*Fig. 2b*). Individual variants can then be created separately at a higher resolution. The time required for the process was about ten minutes.



Fig. 3a. The result of the “crayfish” entry – the first MidJourney attempts



Figs. 3b, c. The final result of entering “crayfish” with the MidJourney tool



Aquatic animal

We tested the creativity and capabilities of the MidJourney AI tool on the creation of depictions of living organisms. Using text input, we let the tool draw a crayfish (Fig. 3a). It turns out that the MidJourney tool generates crayfish with difficulty – adding the wrong anatomy to them. Compiling the prompt required about 10 attempts. Example of a failed prompt:

„A captivating, hyper-realistic underwater photograph of a crayfish with two antennae, gracefully navigating the crystal-clear waters of a mountain creek, showcasing the intricate details and beauty of this fascinating aquatic creature. This stunning image is skillfully captured using a Nikon D850 DSLR camera, equipped with a NIKKOR AF-S 105mm f/2.8G IF-ED VR Micro lens, renowned for its exceptional sharpness and ability to render vivid, life-like colors, even in challenging underwater environments. The camera settings are meticulously chosen to highlight the delicate features of the crayfish and the serene ambiance of its habitat, with an aperture of f/11, ISO 800, and a shutter speed of 1/125 sec. The composition is taken from a close perspective, immersing the viewer in the aquatic world of the crayfish as it scuttles among the rocks and submerged plants that line the creek bed. The scene is softly illuminated by natural sunlight filtering through the water’s surface, casting shimmering patterns that dance across the crayfish’s intricate exoskeleton and the surrounding environment. This awe-inspiring, high-resolution photograph transports viewers beneath the surface of the mountain creek, offering a rare and privileged glimpse into the secret underwater realm of the crayfish. --ar 4:3 --q 2 --v 5.”

After this “failure”, a simple prompt was eventually used: *“A crayfish, captivating, hyper-realistic photograph --ar 4:3 --q 2 --v 5”*. By comparing the first, extensive assignment and the final form, it clearly demonstrates the saying that sometimes less means more :) (Figs. 3b, c). The time required for the process was approximately one hour.

TGM WRI building

The last example on which we tested the AI capabilities was the task of visualizing the building of TGM WRI Prague headquarters, not only in real form (Figs. 4a, b), but also in “Lego” form (Figs. 4c, d). The template was a photo of the TGM WRI building, which was uploaded to the AI tool with the “image to text” command. Some of the elements from the “image to text” description were used and supplemented with a description of lighting, photographic and artistic styles, and colours. The resulting prompt *“a large red and white brick building, in the style of agfa vista, dark bronze and blue, vray, school of london, computer-aided manufacturing, dark brown and navy, lively and energetic -- ar 31:22 -- v 5”* then produced the following result.



Fig. 4a. TGM WRI building (Photo: TGM WRI archive)



Fig. 4b. TGM WRI (visualization using the MidJourney tool)



Figs. 4c, d. TGM WRI building in Lego style – preview image of variants and visualization using the MidJourney tool



Creating the TGM WRI building from Lego bricks required modifying the prompt to the following form: *“a large red and white brick building, in the style of agfa vista, dark bronze and blue, vray, school of london, computer-aided manufacturing, dark brown and navy, lively and energetic, as lego. --ar 31:22 --v 5”*. The result was the generation of a preview image (Fig. 4c). Individual variants can then be created separately at a higher resolution (Fig. 4d). The time required for the process was about 15 minutes.

Conclusion

The MidJourney tool can generate some really nice images, even a bit kitschy in some cases. However, the problem turned out to be that the artificial intelligence does not know what exactly is in the photo. Although it recognizes objects (you tell it to), it cannot assess whether the created image is in accordance with our perceived reality. An example can be the visualization of the font (in our case, the name of our institution on the facade of the building generated by the tool), when the AI tool is not yet able to take the font/signs as parameters from queries. However, Stable Diffusion can already deal with texts.

Due to the relatively dynamic development in the field of artificial intelligence, the functionality and quality of the output in AI applications are constantly changing. For example, the current version of MidJourney already generates very realistic high-resolution images with many details compared to previous versions. On the other hand, there is no detailed documentation of the model on which MidJourney runs, so the resulting graphical outputs vary depending on the form of the prompt that users “fine-tune” based on their experience with the tool, and thus, by “reverse engineering”, they discover possibilities and hidden model settings. To create such a prompt, other AIs are widely used in the form of web applications, which allow the creation of a prompt “tailored” to the desired idea of the output. For this purpose, e.g. ChatGPT will also serve very well.

In particular, MidJourney now has not only the function of creating images, but can also describe others in text after inserting them into the tool and offer its own version. Several image inputs can also be mixed in it and the result is then a composite. It also has numerous choices of styles in which it generates graphics (from imitations of the styles of various artists to animated and anime outputs to photorealistic graphics, e.g. in a fantasy environment). It also allows you to vary the outputs offered almost arbitrarily.

It should be noted that the use of this tool is currently charged and requires registration and login via the Discord service.

References

[1] HRDINKA, T. Plánované revitalizace vodních ekosystémů v Praze 4. *Vodohospodářské technicko-ekonomické informace*. 2023, 65(2), pp. 52. Available at: <https://www.vtei.cz/2023/04/planovane-revitalizace-vodnich-ekosystemu-v-praze-4/>

[2] KORÍNEK, R., HORÁČEK, M. Představení některých výsledků výzkumného projektu mapující vývoj staveb věžových vodojemů na území České republiky. *Vodohospodářské technicko-ekonomické informace*. 2022, 64(6), pp. 8. Available at: <https://www.vtei.cz/2022/12/predstaveni-nekterych-vysledku-vyzkumneho-projektu-mapujici-vyvoj-staveb-vezovych-vodojemu-na-uzemi-ceske-republiky-2/>

Authors

Ing. Josef Nistler

✉ josef.nistler@vuv.cz

Tomáš Pojeta

✉ tomas.pojeta@vuv.cz

T. G. Masaryk Water Research Institute, Prague