Artificial intelligence vs human designers: partners or competitors

Inteligencia artificial versus diseñadores humanos: socios o competidores

Anastasiia Dovgal [1]
[1]co-owner, art-director of design studio Glad Head, Kharkiv, Ukraine
Yuliia.borovynska@nure.ua[1]

Abstract

The study explores the processes of generating images using artificial intelligence products DALL-E 2 and Midjourney with the aim of understanding the impact of Artificial Intelligence on designers' work. To reinforce the conclusions, a series of images were generated across various prompts in Midjourney).

Keywords: artificial intelligence, DALL-E 2, Midjourney, design, creation, generate, image, CLIP, GLIDE, diffusion, algorithm.

Introduction

Artificial intelligence (AI) is increasingly becoming an integral part of the lives of modern designers. Currently, with the help of artificial intelligence, it is possible to significantly simplify and reduce the time spent on routine design processes such as:

- brainstorming;
- overall process organization;
- photo retouching;
- color palette selection;
- applying styles and filters;
- prototyping;
- creating wareframes;
- layout;
- creating icons;
- simple illustrations, etc.

This list is continuously expanding and evolving, reflecting the dynamic nature of design practices. Designers from diverse backgrounds and industries are increasingly embracing the transformative capabilities of AI in their daily workflows. With AI becoming an integral part of the design process, professionals are leveraging its power to unlock new levels of creativity, efficiency, and innovation. As AI technologies advance, designers are boldly exploring its potential across various spheres, revolutionizing the way design is conceptualized, executed, and experienced.

We can delve into entirely novel and intricate processes, such as real-time image generation, which emerged just moments ago. Additionally, we can explore the substantial streamlining of well-known workflows achieved by integrating artificial intelligence seamlessly into familiar software platforms like Adobe and Figma. This integration marks a paradigm shift in design methodologies, empowering designers to innovate and create with unprecedented efficiency and creativity.

Today, the use of artificial intelligence is not only the domain of researchers, scientists, or behind-the-scenes optimization processes. More and more well-known companies are utilizing AI in their advertising campaigns and promotions. Heinz ketchup is about to launch an advertising campaign where all the images are generated using artificial intelligence (fig. 1), where in different styles the famous ketchup can is depicted. Cosmopolitan has released a cover (fig. 2) that was created only using artificial intelligence. Coca Cola has launched an advertising campaign called "Create Real Magic" (fig. 3), where you can create your own advertisement for the drink.

Figure 1. Heinz ketchup advertisement generated using Dell-E 2 artificial intelligence, 2023.

Figure 2. New York, USA 22 June 2022 - Cosmopolitan is one of the first print publications to leverage OpenAI's cutting-edge image generator, Dall-E 2, to help design its cover art for its latest issue.

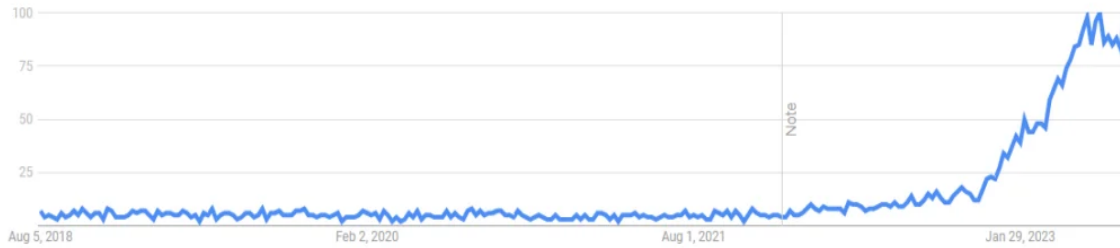Figure 3. "Create Real Magic" is the first output of The Coca-Cola Company's collaboration with AI, 2023.

Figure 4. According to Google trends, the search volume for AI design related tools and software has increased 1700% from 2022 to 2023 alone.

According to recent research findings, the utilization of artificial intelligence in graphic design has surged by a staggering 1700 percent over the past year (fig. 4), (Marq.com., 2023). Such a substantial growth rate is undeniably impressive! It's becoming increasingly apparent that many designers are starting to express concerns about the potential for artificial intelligence to completely replace human labor in the field. So, will AI be able to completely replace the human designer, or will it become a reliable assistant that will expand the horizon of possibilities for human designers?

Methodology

At present, there is no possibility to answer this question with complete certainty since AI surprises us every day. In the field of graphic design, the most amazing ones at the moment are the Midjourney and DALL-E 2 Image Generators. These platforms are capable of creating incredibly beautiful and realistic images, illustrations or even paintings in an amazingly short time, using only a short written request. Let's look at the AI process using these two platforms as an example.

The process of creating an image in DALL-E 2.

At the time of writing the article, the DALL-E 3 version is already available, but since it is still poorly studied, in this article, we will describe DALL-E 2.

1. Input Text. The first neural network is called CLIP, it translates our written (human) text into computer language in the form of numbers by coder Word2Vec (DALL-E 3 additionally uses GPT-4).

2. Text Encoding. Next, CLIP converts this set of numbers into a table with other numbers. Such a table serves as a "draft" for creating the final image. To make it all work, CLIP was trained on 600 million pictures and their captions.

3. Image Generation. The "draft" passes to the second neural network called GLIDE. GLIDE takes the initial computer text from step 1 and the resulting scheme from step 2, combines the data from them. Based on this mix, it creates a gray grainy square, from which it gradually removes the grain, thereby revealing the image in poor quality. This development method is called "Diffusion model application."* The Diffusion model takes textual and visual vectors and outputs an image with a resolution of 64 x 64.

4. Evaluation. The final image is scaled to 1024 x 1024 pixels using two additional diffusion models.

DALL·E 2 isn't solely a neural network but rather a composite product amalgamated from various programs, often termed as a generative model within professional discourse. It leverages the functionalities of two preceding products developed by OpenAI: CLIP and GLIDE. Eventually, a third neural network is deployed to enhance the image resolution. The incorporation predominantly involves specific components from CLIP and GLIDE, rather than the complete programs (fig. 5). The method of contextual comprehension was adopted from the former, while the latter contributed its superior visualization quality (Novrapture, 2024, Ramesh A., 2023).
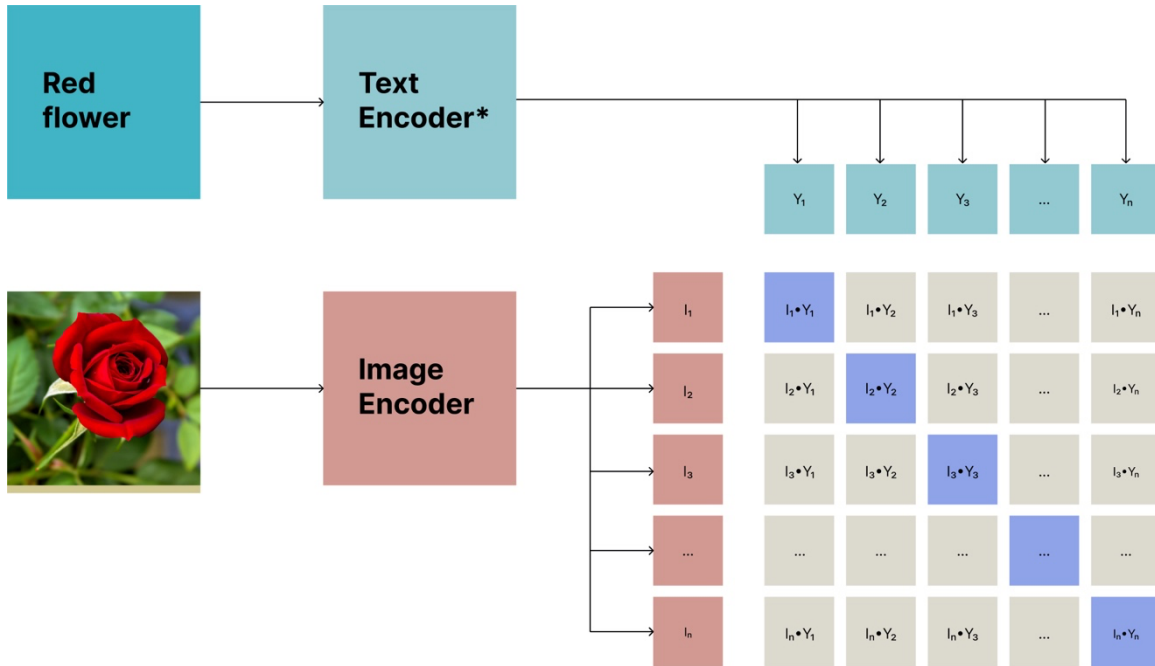
Figure 5. Illustration of the contrastive training objective for CLIP.

During each step of training, CLIP receives X=32,786 images and their corresponding captions. From these, we form X matching image-caption pairs (corresponding to the diagonal elements of the matrix in the illustration), and $X(X-1)$ pairs of mismatching captions and images (corresponding to the off-diagonal elements

DALL·E 2 isn't solely a neural network but rather a composite product amalgamated from various programs, often termed as a generative model within professional discourse. It leverages the functionalities of two preceding products developed by OpenAI: CLIP and GLIDE. Eventually, a third neural network is deployed to enhance the image resolution. The incorporation predominantly involves specific components from CLIP and GLIDE, rather than the complete programs. The method of contextual comprehension was adopted from the former, while the latter contributed its superior visualization quality.

Imagine a small boat on the blue sea. The way you pictured the scene is the result of CLIP's work. You may feel the main details of this vision, but you may not immediately recall or imagine the finer details. CLIP converts text into "text embedding" - a digital description of words. This is not just encoding words like binary code, but rather representing words as tokens and vectors according to their meaning. During training, these tokens and vectors are classified into more than 300 categories, the exact composition and ranking of which are unknown to developers because CLIP forms them itself. And GLIDE converts text into an image. It's like DALL-E 2, but with fewer steps and less accuracy. Essentially, this neural network could be called DALL-E 1.5 if OpenAI gave loud names to each of its generative models. The working method of GLIDE provides the same realism that shocks everyone. But not because it's smart, but because it already has a "skeleton" created by CLIP algorithms.

The process of creating an image in Midjourney.

1. Input Processing. Midjourney first preprocesses and tokenizes the input text to extract relevant information. This involves removing stop words, punctuation, and other noise from the input text.

2. Text Understanding. Natural Language Processing (NLP) techniques are used to understand the semantics and context of the input text. This includes tasks such as named entity recognition, sentiment analysis, and semantic parsing to extract key concepts, attributes, and relationships.

3. Image Generation Model. Midjourney likely employs a deep learning architecture, such as a generative adversarial network (GAN) or a transformer-based model, to generate images from the processed textual

descriptions. These models are trained on large datasets of paired text-image examples to learn the mapping between text and images.

4. Attention Mechanisms. Transformer-based models, commonly used in modern AI architectures, utilize attention mechanisms to focus on relevant parts of the input text during the image generation process. This helps the model capture fine-grained details and nuances from the textual descriptions.

5. Fine-tuning and Iteration. Midjourney may incorporate fine-tuning techniques and iterative refinement processes to enhance the quality and realism of the generated images. This involves adjusting model parameters, optimizing loss functions, and incorporating feedback from users to iteratively improve the image generation process.

6. Output Optimization. The final output of Midjourney is an image that visually represents the concepts and elements described in the input text. Post-processing techniques such as image resizing, color adjustment, and style transfer may be applied to enhance the visual quality and coherence of the generated images.

Despite DALL-E2 and Midjourney using slightly different platforms in the image generation process, it's evident that overall, they have a very similar mechanism.

As we can see, they are trained to mimic human artistic abilities. These generators "learn" specific styles or aesthetics by analyzing vast datasets containing thousands to millions of images. Understanding the relationship between visual information and corresponding textual descriptions, such systems are capable of generating their own images in response to textual prompts. No matter how shocking the speed and quality with which AI creates new images, the starting point remains the command of a human.

Artificial intelligence simulates human thinking through computer systems. These systems, trained to analyze information, can perform a wide range of tasks similar to, but importantly not identical to, those performed by humans. Perhaps everything will change in the very near future, but as of today, the main difference between the work process of artificial intelligence and that of humans is that AI cannot evaluate your creativity.

Currently, creating professional work without human involvement is not possible (Yup K., 2003). Any AI resource requires a reviewer. No matter how high-quality and fast the AI works, to implement these developments into the final design, there always needs to be someone who will evaluate the AI's work and then incorporate it into the project. AI helps with analysis and selection, but it cannot evaluate its own work, find and correct errors, or improve it in any way. AI executes commands, and someone must give it those commands. Evaluation and criticism of the AI's work will also not yield good results if it is not done by a professional, someone without artistic education, without experience in graphic design, without experience in managing business processes, and understanding the target audience.

Below is an example of creating an image using Midjourney. The goal was to create a high-quality, aesthetic image (fig. 6).

Figure 6. Shot 1 / Promt: Darth Vader is a DJ.

The first attempt turned out decently, but I wanted to add more details and a different style (fig. 7).

Figure 7. Shot 2 / Promt: Darth Vader is a DJ, Darth Vader close-up, warriors of the empire and inhabitants of different planets dancing around, the action takes place inside a spaceship, detailed, aesthetic, realistic - ar 4:5.

Now I want to add more cinematography and realism to the image (fig. 8).

Figure 8. Shot 3 / Promt: Darth vader is a DJ, he is strict and stern, warriors of the empire and inhabitants of different planets dancing around, Darth Vader takes 2/3 part of the picture, black & white, minimal, photorealistic, cinematic, detailed, aesthetic, realistic --ar 4:5.

The quality and style are excellent, but somehow all the surroundings disappeared, and Art Vader is no longer a DJ. My goal was to show that he is a DJ and plays in the club, adding equipment, dancing people, and the appropriate atmosphere (fig. 9-11).

Figure 9-11. Several more attemps.

There were several more attempts, until finally, with the addition of specific descriptions and parameters, I got exactly what I wanted (fig.12)

Figure 12. Shot N / Promt: Darth vader is a DJ, he is strict and stern, the crowd dance around, black & white, minimal, photorealistic, cinematic, detailed, aesthetic, realistic:: --aspect 16:9 --version 6.0 --stylize 20 --chaos 1 --weird 0.

Results and Conclusions

Thus, it can be concluded that AI will be an irreplaceable assistant for designers of the future and that modern designers must incorporate AI into their daily work to keep up with the times, reduce time spent on monotonous processes, and spend it on creating something truly creative and outstanding. Integrating AI into design workflows will not only enhance efficiency but also foster innovation and creativity in the field.

As AI continues to evolve, designers should view it as a tool to augment, rather than replace, human creativity. Collaborations between human designers and AI systems hold the potential to unlock new realms of possibility in design. The symbiotic relationship between AI and live designers promises to shape the future of any kind of design. By leveraging the power of AI as a complementary tool, designers can navigate the changing design landscape, adapt specific needs and challenges to new technologies, and continue to create innovative designs much faster and more efficiently.

Explanations

Diffusion model – a method of manifesting images from pure noise, used by the decoder.

Encoder – a neural network that encodes human text and images into a language understandable by another neural network in the form of vectors.

Image embedding – a set of numbers that describes each word or part of an image. Each neural network has its own.

Decoder – a neural network that converts vectors into human-understandable language or images.

Transformer – a common term for encoders and decoders.

References

Rethink. (2023). https://www.adsoftheworld.com/campaigns/a-i-ketchup.

Cosmopolitan. (2022). https://www.cosmopolitan.com/lifestyle/a40314356/dall-e-2-artificial-intelligence-cover/.

Coca-Cola. (2023). https://www.coca-colacompany.com/media-center/coca-cola-invites-digital-artists-to-create-real-magic-using-new-ai-platform

Marq.com. (2023). https://www.marq.com/blog/artificial-intelligence-design-tool-statistics-trends-in-2023#:~:text=limited%20to%20design.-,The%20Rising%20Demand%20for%20AI%20Design%20Tools,from%202022%20to%20 2023%20alone.

Novrapture. (2024). https://www.iphones.ru/iNotes/kak-rabotaet-neyroset-dalle-2-obyasnyayu-na-tryoh-urovnyah-ponimaniya-08-18-2022#pic1153877

Ramesh, A. (2023). http://adityaramesh.com/posts/dalle2/dalle2.html.

Yup, K. (2003). https://yaledailynews.com/blog/2023/01/23/what-ai-art-means-for-society-according-to-yale-experts/.