A COMPARISON OF ARTIFICIAL INTELLIGENCE IMAGE GENERATION TOOLS IN PRODUCT DESIGN

Sam DHAMI and Ross BRISCO
University of Strathclyde, United Kingdom

ABSTRACT
Artificial intelligence (AI) image generators have seen a significant increase in sophistication and public accessibility in recent years, capable of creating photorealistic and complex images from a line of text. A potential application for these image generators is in the concept generation phase in product design projects. Successful implementation of AI text-to-image generators in concept generation could prove to be a cost and time saving application for companies and designers. Therefore, the aim of this paper is to investigate the integration of AI into product design and education. A literature review was conducted to gain a general understanding of what AI is and how AI image generators function. An experiment was carried out which used three different image generators: Stable Diffusion, DALL·E 2, and Midjourney. Three images of dining tables were produced by each AI text-to-image generator and inserted into a weighting and rating matrix to be rated as concepts along with three real dining tables from IKEA. Within the matrix were four design specifications to rate the concepts against: aesthetics; performance; size; safety. The matrix was sent out to product design students and graduates to be completed anonymously. The highest scoring concept was one from IKEA, followed by one generated by DALL·E 2. Based on the results of the experiment, it was concluded that AI image generators are not yet a viable alternative for concept generation in product design but could be a useful tool to spark new ideas for designers to use during the concept generation phase.

Keywords: Artificial intelligence, product design, text-to-image generation, concept generation

1 INTRODUCTION
In product design, a project will typically require four stages: initial research, concept generation, concept evaluation & selection, and detailed design. Artificial Intelligence (AI) has the potential to support designers at each of these four stages. AI image generators is one promising tool which could support concept generation. These tools are capable of creating photorealistic and complex images from a line of text. However, the suitability of this technology to support designers is unclear. In this paper, the suitability of using AI image generators for the concept generation stage in student's product design projects is investigated. Although the incorporation of AI in product design is an emerging research area, successful implementation of AI text-to-image generators in the concept generation phase could prove to be a cost and time saving application for companies, designers, and design students. If educators have a better understanding of the value of these tools, and future successful and unsuccessful use cases then they can best advise students and educate them towards successful integration of these tools in the design process.

2 LITERATURE REVIEW
AI as a concept has been researched and developed since the late 1940's with its genesis attributed to Alan Turing who posed the question "Can machines think?" [1] and John McCarthy who invented the term 'Artificial Intelligence' and defined it as "the science and engineering of making intelligent machines, especially intelligent computer programs" [2]. In general, AI can be interpreted as creating a machine which is able to operate intelligently or even be able to think by itself, just as a human can.

2.1 AI Text-to-Image Generation
AI text-to-image generators have seen a dramatic increase in sophistication and public availability in the last few years. Simply, they are programs/applications that present a prompt to enter a text description of an image and return an image that is accurate to the text description. There are multiple

methods that can be used to generate images; however, this paper will focus on popular methods of generative adversarial networks (GANs) and diffusion. This explanation is aimed to support the wider understanding of the types of technology used and how this works when using the label of AI.

2.1.1 Generative Adversarial Networks

A GAN is comprised of two neural networks: a generator and a discriminator. The generator takes random gaussian noise and produces a fake image that is fed into the discriminator along with a real image of the same subject. The discriminator tries to identify which image is real, whilst the generator tries to fool the discriminator by making the fake image look real [3]. To begin, the generator may create an image where the discriminator has no trouble in identifying the real image. The generator uses the discriminator's identification and calculates a loss (generator loss) from it, where it undergoes backward propagation through both the discriminator and generator. The weights of the generator are updated, and this concludes one iteration of training for the generator [4]. If the discriminator misidentifies the real image as the fake and the fake image as real, the weights of the discriminator are adjusted based on the discriminator loss which is fed back through the discriminator network [4], concluding one iteration of training for the discriminator. The aim for training the GAN is to get to a point where the generator is able to consistently produce images where the discriminator cannot distinguish between the fake and real images. When the generator is fully trained, word embedding is used to ensure the generator represents the words used accurately. Word embedding is the process of training a network to recognise the meaning of a word and its relation to other words of a similar meaning [5] so the network knows what to use to represent these words.

Although GANs are capable of producing photorealistic images, training these networks can be very difficult because of a problem known as 'mode collapse'. This is when the generator creates the same image (with minimal changes) because it has worked out how to beat the discriminator and, therefore, has no incentive for creating interesting new images. It has found a solution for the problem it was given so it sees no need in finding other solutions for the same problem. Moreover, producing a photorealistic image with no blemishes from random noise is no easy task. GANs can suffer from obscurities and errors in their images when it comes to generating complex images. Diffusion can solve these issues.

2.1.2 Diffusion

Where GANs generate images directly from random noise, diffusion models use a neural network that takes this process in iterative steps. Training a diffusion model involves taking an image and gradually adding gaussian noise to the image until it is complete random noise (forward diffusion). This process is then reversed, where the neural network tries to discern how much noise was added by calculating probabilities at each stage in order to recover the original image (denoising). This process is known as reverse diffusion. Once the diffusion model learns the reverse diffusion process and can denoise images accurately, it can be used to generate new images from random noise inputs [6].

Contrastive Language-Image Pretraining (CLIP) is a tool used in conjunction with the diffusion process to fine-tune images and make them as accurate to the text description as possible. It uses an image and a text encoder that have been trained to find which texts correlate to certain images within a dataset as described by Kim, Kwon and Ye; "The input image is first converted to the latent via diffusion models. Then, guided by directional CLIP loss, the diffusion model is fine-tuned, and the updated sample is generated during reverse diffusion." In this context, the "latent" is the random noise that the image from the dataset is converted to in the forward diffusion process. Using the loss function from CLIP the diffusion process is adjusted accordingly, and an optimised image is produced from reverse diffusion which should be more accurate to the initial image request [7].

3 METHODOLOGIES

To investigate whether AI text-to-image generators would be viable for generating concepts for student's product design projects, an experiment was created using three AI image generators to each produce three images of a dining table. These images were rated by participants as 'product concepts' using a rating matrix. Pugh's elements of specifications were used to establish generic design requirements. Four specifications were selected being: aesthetics; performance; safety and size.

Participants were 5th year students and graduates of the department of Design, Manufacturing and Engineering Management (DMEM) at the University of Strathclyde with experience from studying product design and with familiarity of the rating matrix used to collect data. The rating matrix was created digitally using a survey tool (Qualtrics) and was emailed to participants. 12 participants took part in the study and were of ages between 21 and 24.

Figure 1. Image examples (from left to right) Ikea, Stable Diffusion, DALL·E 2, and Midjourney

Participants were not made aware that three images of real dining tables available from IKEA were included in the evaluation with 9 concepts generated by AI. If the AI text-to-image generated images scored higher than the real tables, the AI will have succeeded in creating viable concepts for product design, otherwise, it will have failed. The AI text-to-image generators used for this experiment were Stable Diffusion, DALL·E 2, and Midjourney. Images of dining tables were selected from the IKEA website. Images from Stable Diffusion, DALL·E 2 and Midjourney were created in November 2022 using the text prompt "a dining table". Where AI tools generated multiple images, the first three presented were selected.

4 RESULTS

The sum of the ratings of the participants was calculated for each table and the results are presented in Figure 2. Tables 12, 8 and 3 are those created by Midjourney (Blue), Tables 10, 6 and 5 were created by DALL·E 2 (Green), Tables 9, 7 and 2 were created by Stable Diffusion (Orange), and Tables 11,4 and 1 are taken from the Ikea website (Yellow).

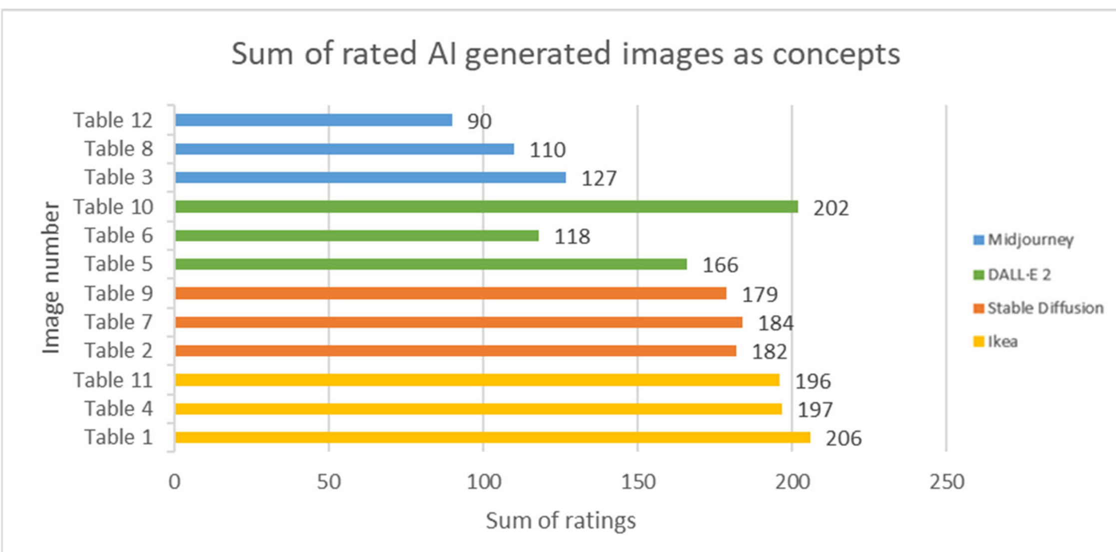

Figure 2. Sum of rated AI generated images as dining table concepts

The concept with the highest score of 206 was Table 1 taken from the IKEA website. The lowest scoring table with a score of 90 was Table 12 generated by Midjourney. All 3 Midjourney tables were among the 4 lowest rated. The most consistently highest scoring set of tables were those generated by Stable Diffusion, with only a 4.5-point difference between the highest and lowest rated tables. The key result from this experiment was DALL·E 2's Table 10 which rated higher than two of the IKEA tables with a score of 202.

5 DISCUSSIONS

Before discussing the results of the experiment, the understanding of what a concept is to the participants should be established. Our students are taught that a concept is a visual representation of an idea. These ideas can be expressed through various mediums such as 2D and 3D sketches, computer renders and physical prototypes. We teach our students to minimise the constraints on the generation of initial concept generation to support ideation. Only when evaluating concepts would they typically begin to introduce constraints in the form of product design specifications.

AI text to image generation tools do not currently have the ability to generate concepts based on design specifications which support creative ideation. However, where they are limited is in the data set used to train them. This may not support a wide exploration of ideas however there is evidence that data-driven design concept generation approaches have great value in supporting novice designers when incorporate novel technologies [8].

Beyond this, it should be noted that the ratings provided by the respondents are subject to the interpretation of the image as a concept. Each product designer will have their own interpretation of what a dining table should look like, the features and functionality the concept should have, and the same can be assumed for each of the design requirements used to evaluate the tables within the rating matrix. This may be a limitation of the study to not provide a design brief to the respondents, however, this may be authentic to the real-life process of evaluating and selecting concepts for some product design projects.

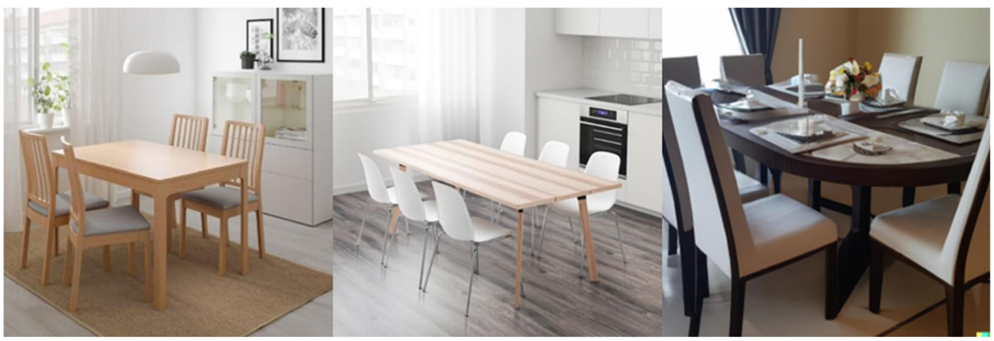

Figure 3. (from left to right) Table 4, 11 (Ikea) and 10 (DALL·E 2)

Analysing the results of the experiment, the AI has both succeeded and failed in scoring higher than the tables from IKEA. The outcomes of the images that AI created are inconsistent as concepts. Although DALL·E's Table 10 scored higher than IKEA Tables 4 and 11 (Figure 3), this should not be misconstrued as DALL·E 2 being the viable option for generating concepts suitable for product design. Table 10 is surrounded by chairs which mask the table's legs and makes it difficult to discern the chair legs from the table. This was a limitation of the AI, the ambiguity of the image concealing the table's legs has likely played to the interpretation of the concept.

Figure 4. Images generated by Midjourney (Table 3, 8 and 12)

Dining tables generated by Midjourney (Figure 4) were among the lowest rated of all twelve tables. When considering the images generated comparing them to Stable Diffusion and DALL·E 2 it is clear why Midjourney had a lower score. The images are more artistic and abstract in nature and appear as though they have been drawn or painted. As a result, the dining tables generated made little sense in terms of the design requirements in the matrix. The main reason for this is most likely due to Midjourney's training data consisting of drawings, paintings, and other artistic works. Whereas Stable Diffusion and DALL·E 2 were, presumably, trained more with photographs which is why their images generated look photorealistic instead of hand drawn. This raises the question about the appropriateness of the training data to the use case of the AI text-to-image generator. If an AI text-to-image generator were created with product images, would it be appropriate and more successful for the purpose of evaluating and ultimately selecting concepts? Similarly, if a training set was created of a single product category would it be able to generate novel conceptual designs?

It may also be assumed that Stable Diffusion and DALL·E 2 have more images of dining tables within their training data compared to Midjourney which may have added to the odd and abstract look of the dining tables. Although detrimental to the outcomes of the experiment, Midjourney producing concepts through an artistic medium could be beneficial to some product designers since they would typically sketch initial concepts in an abstract way. If the training data was tailored more towards various objects and products that would be suitable for product design, then Midjourney could be a viable option for providing inspiration for new and innovative concepts.

Figure 5. Images generated by Stable Diffusion (Table 2, 7 and 9)

Stable Diffusion produced the most consistent high scoring images of dining tables (Figure 5), with only a 4.5-point difference between its highest and lowest rated tables. Compared to DALL·E 2 which had an 84-point difference and Midjourney which had a 37-point difference. Also, Stable Diffusion's Table 7 was 22.5-points lower than IKEA's Table 1 which was the highest scoring dining table. This may be another reflection of the impact that the training data can have on the outputs from the AI text-to-image generators.

5.1 Reflections for educators

With the rate that acceptance of AI is growing, and how quickly it is advancing, the capabilities of AI image generation may soon be at a stage where it could become a common tool used by product designers to help with concept generation or even be the forefront of it in a range of roles [9]. Moreover, further advances in AI will allow for even more complex and sophisticated machine learning models to be created. An example could be a deep learning model built specifically with product design in mind. One that is incorporated with computer-aided design (CAD) and can easily create a wide range of concepts that take design specifications and engineering parameters into consideration with minimal input from designers. It could become common practice to use this model for the entire concept generation phase, saving designers a significant amount of time on the project. As new research on using AI in a product design environment begins to emerge, the likelihood of a deep learning model such as this becoming a reality increases.

If such a system were to exist, there may come a time where some everyday products have solely originated from artificial intelligence. As educators we cannot stop our students from experimenting with these tools and we should not discourage it. Once the capabilities of the tools catch up with the use case to generate product design concepts, our students need to understand how best to use these tools and we can begin to build their competencies in forming text prompts for these systems. As researchers we can support the understanding of the capabilities of current tools and drive the development of tailored tools to deliver solutions that are more accurate to the task. We should not encourage the use of novel technology without reason [10]. However, we can have a hand in supporting the evaluation of possible technologies from our investigations into their affordances.

6 CONCLUSIONS

This paper aimed to investigate if AI could be integrated into student's product design process using text-to-image generation during the concept generation phase. An initial research investigation determined that the text-to-image generation method utilised, the training method and the interpretation of the images all play a significant role in the suitability of generated images as product design concepts. Tools such as DALL·E 2 were able to generate images that outperformed real world images which indicates a potential for this technology. However, the majority of concepts performed lower in a concept evaluation activity. There are still questions to be asked in the role that specifications and interpretation play in evaluating generated images. And if these challenges can be overcome with specific AI tools are created in the future which focus on inspiration of ideas or specific product categories.

Most text to image generation tools currently available are free to use, easily accessed, easy to use, and capable of generating images almost instantaneously, there are few reasons for designers within industry and education not to experiment with them, despite current limitations. As educators, we should communicate how best these tools can be used and when not to use them. For now, these tools have a place as inspiration when the design student has run out of ideas. However, currently they cannot replace the creativity of the designer, nor the intuitiveness of their response to design challenges within the conceptual generation phase of the design process.

REFERENCES

[1] Turing A. M. (1950). Computing Machinery and Intelligence. Mind, LIX (236), pp.433–460. doi: 10.1093/mind/lix.236.433.

[2] McCarthy J. (2007). What is AI? Stanford University. Available at: http://jmc.stanford.edu/artificial-intelligence/what-is-ai/index.html.

[3] Mao X. and Li Q. (2020). Generative Adversarial Networks (GANs). In: *Generative Adversarial Networks for Image Generation*. Springer Singapore, pp.1–7. doi: 10.1007/978-981-33-6048-8_1.

[4] Google Developers (2022). Overview of GAN Structure. *Machine Learning Advanced Courses.* Available at: https://developers.google.com/machine-learning/gan/gan_structure.

[5] Goldberg Y. (2017). Pre-trained Word Representations. In: *Neural Network Methods for Natural Language Processing*. Cham: Springer International Publishing, pp.115–134. doi: 10.1007/978-3-031-02165-7.

[6] Song Y., Sohl-Dickstein J., Kingma D. P., Kumar A., Ermon S. and Poole B. (2021). Score-Based Generative Modeling Through Stochastic Differential Equations. In: *The International Conference on Learning Representations 2021*. ICLR 2021. doi: 10.48550/arXiv.2011.13456.

[7] Kim G., Kwon T. and Ye J. C. (2022). DiffusionCLIP: Text-Guided Diffusion Models for Robust Image Manipulation. In: *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. pp.2416–2425. doi: 10.1109/CVPR52688.2022.00246.

[8] Han J., Forbes H., Shi F., Hao J. and Schaefer D. (2020) A data-driven approach for creative concept generation and evaluation, *Proceedings of the Design Society: DESIGN Conference.* Cambridge University Press, 1, pp. 167–176. doi: 10.1017/dsd.2020.5.

[9] McComb C., Boatwright P. and Cagan J. (2023). Focus and modality: defining a roadmap to future ai-human teaming in design. *Proceedings of the Design Society*, 3, 1905–1914. doi: 10.1017/pds.2023.191.

[10] Beetham H. and Sharpe R. (2019). *Rethinking pedagogy for a digital age: principles and practices of design.* Routledge. doi: 10.4324/9781351252805.