RESEARCH PLAN

ML8 – A TRIPLE OF IMAGE GENERATION TOOLS

Group name: Left4Dead

List of members:
Nguyễn Thế Thiện - 21127170
Châu Tấn Kiệt – 21127329
Trịnh Minh Long - 21127642

Keywords: image generation, DALL-E, Imagen, Midjourney, machine learning

Description: This research is about 3 image generation tools namely DALL-E, Imagen and Midjourney: an in-depth analysis on the scientific basis of the ideas, their strengths and weaknesses, best use cases and real-life implementation of the tools.

List of references:

- [1] ACM SIGGRAPH, Harold Cohen (2014, August 22). ACM SIGGRAPH Awards - Harold Cohen, Distinguished Artist Award for Lifetime Achievement. YouTube: https://www.youtube.com/watch?v=_Xbt8lzWxIQ.
- [2] Karl Sims (1992). Liquid Selves. *Art Futura and "Memory Palace" at the 1992 World's Fair in Spain*. http://www.karlsims.com/liquid-selves.html
- [3] Will Knight (2023, Jan 12). Where the AI Art Boom Came From—and Where It's Going. *WIRED*. https://www.wired.com/gallery/where-the-ai-art-boom-came-from-and-where-its-going/.
- [4] J. Becker, G. Goh, Li Jing et. al. (2023). Improving Image Generation with Better Captions. *OpenAI*. https://cdn.openai.com/papers/dall-e-3.pdf.
- [5] C. Saharia, W. Chan, S. Saxena et. Al. (2022, May 23). Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding. *Canada: Google Research, Brain Team.* ArXiv https://arxiv.org/pdf/2205.11487.pdf
- [6] Matthew Gault (2022, Aug 31). An AI-Generated Artwork Won First Place at a State Fair Fine Arts Competition, and Artists Are Pissed. *VICE*. https://www.vice.com/en/article/bvmvqm/an-ai-generated-artwork-won-first-place-at-a-state-fair-fine-arts-competition-and-artists-are-pissed.
- [7] Everypixel Journal (2023). AI Has Already Created As Many Images As Photographers Have Taken in 150 Years, Statistics for 2023.

https://journal.everypixel.com/ai-image-statistics. Estimation by Adobe, Photutorial, OpenAI, Civitai (2023, Aug).

Presentation outline

1. **Introduction:** an overview of the selected solution (i.e., a tool, a project or a product that provides great helps to Big Data research)

   - Which problems could be solved by using this solution?

   In Big Data, images are a way to visualize data in the graphical form, and in different points of view could we extract distinct facts, knowledge or any potential analysis directions. These image generation tools help picture the ideas in text (prompts) into images, or create variants of an existing image, which could solve lack in data needed for additional accuracies in data analysis.

   - A brief history: motivations and evolution through years

   The concept of "computer programs that create original artistic images" by Harold Cohen started in the late 1960s, as the computer program series AARON was developed from only able to draw in black and white at first to get more colors and more imagery shapes, even though could not learn new styles and had to be hand-coded [1].

   The 1990s and 2000s went by with integrations of artificial evolution into image processing with notable arts such as Liquid Selves [2], but with not much progress in this era. The boom in Big Data and AI-based image processing actually started in the 2010s with the rise of Deep Learning and Convolutional Neural Networks for high quality photo editing [3], thus the big companies jumping into this potential land to take the spots: DALL-E (OpenAI), Imagen (Google DeepMind), plus a new independent player coming into the game: Midjourney (Midjourney Inc.). The three big names came into media spotlight in 2023.

2. **A deeper insight to the selected solution**

   2.1. Major components and main functionalities: consider what are worth to be presented and support them with demonstrative

February 25, 2024

figures and examples

2.2. Its applications in academic and/or industry activities

2.3. Its popularity (i.e., how many users choose this solution)

2.4. Identify other solutions that have similar functionalities and compare them on different aspects

Tools	DALL-E	Imagen	Midjourney
Main functionalities	Image generation from prompts. Replacing areas of images, generating variants of images. (ver 2 only)	Image generation from prompts. Edit image or part of image using text prompt. Generate captions, extract information from image.	Image generation from prompts. (specialized) Integrated collaborative artwork-sharing platform.
Major components	Generative Pre-trained Transformer (GPT) model Contrastive Language-Image Pre-training (CLIP) model, including zero-shot learning. [4]	Based on Stable Diffusion model. Text encoder, image generator, super resolution models [5].	Midjourney model (closed-source, but similar to Stable Diffusion)
Applications	Specialized in text-to-image generation. Innovative, applied artistic styles in images.	Scientific illustrations. Data visualization. Logo, advertisement design.	Specialized in creating arts and illustrations [6]. Designs for advertisements, marketing and creative arts.
Popularity [7]	916 million images (DALL-E 2 only)	12.59 billion images (altogether for models based on Stable Diffusion)	964 million images

3. Demonstration

- **Main idea**: testing the three tools with the same test prompts/images in different areas, in order to make comparisons and figure out which tool is the best of the three in any field.
- Steps:
1. Short installation/integration guides on how to use the tools.
2. Determine the problem, environment, settings into testing.
3. Test and compare the results based on the use cases and the purposes of different fields in question.
4. Conclusion on choosing which tool to use and what to expect from these tools in the future.

4. Discussions and Conclusion

Weekly schedule: what each member has done every week

	Week 01 (from 18/02 to 24/02)	Week 02 (from…to…)		Week n (from…to…)
Thiện	Tool research: DALL-E Research plan fill-in.			
Kiệt	Tool research: Midjourney.			
Long	Tool research: ImageGen.			

Group self-evaluation: State the objective and subjective reasons that may affect your performance in completing this research; however, it is not about evaluating the contribution of team members.

- **Advantages:**
  - o Imagen, DALL-E are open-source, could look into how they work and the (possible) biases, algorithm weaknesses and strengths, and

new knowledge.

- o Good researchability as the names and topic are very new (2023), and still have more potential to cover.

- **Disadvantages:**
  - o Midjourney is closed-source, the team is still questioning if it is worth diving deeper into the algorithms of the tools by examining the codes.
  - o Lack of correct data about popularity, so far the collected data is estimated from a collection of known sources.
  - o Most known sources from medias covering on this topic do not get deep into the models and technologies as their intended audience is the general public.