AI Image Generator in Digital Illustration Creation: A Literature Review

Elly Herliyani [a,1,*], Ketut Agustini [b,2], I Gde Wawan Sudatha [b,3], Gede Rasben Dantes [b,4], I Gusti Putu Suharta [b,5], I Kadek Suartama [b,6]

[a] Universitas Pendidikan Ganesha, Udayana number 11, Singaraja and 81116, Indonesia
[b] Universitas Pendidikan Ganesha, Udayana number 11, Singaraja and 81116, Indonesia
[1] elly.herliyani@undiksha.ac.id*; [2] ketutagustini@undiksha.ac.id; [3] igdewawans@undiksha.ac.id; [4] rasben.dantes@undiksha.ac.id; [5] putu.suharta@undiksha.ac.id; [6] ik-suartama@undiksha.ac.id

ARTICLE INFO

Article history
Received
Revised
Accepted

Keywords
AI Image Generator
Digital Illustration
Visual Communication Design

ABSTRACT

The use of artificial intelligence image generators in creating digital illustration is experiencing rapid development as an alternative to support the work of designers in various fields, especially visual communication design. Despite its sophistication, we must consider some aspects of the human side in order to maximize the design aesthetic value. This research aims to find out what applications are used in the AI image generator, as well as to identify the results of digital illustrations using the AI image generator. The results of this research show that the image generation diffusion model with various techniques and variations developed can produce excellent and unique text-based quality and control. However, humans are still the best pilots for determining text-based commands that can produce and function the greatest result.

1. Introduction

Creativity today has gone beyond traditional ways of working when using artificial intelligence (AI), thus it influences the aesthetics and technicalities of exploring art in creative practice [1]. Moreover, the creative practice of artists with AI in the field of human-computer interaction can enhance creativity, optimize production processes and engage audiences through interactive installations [2]. Especially with the presence of humanoid robot artists on Ai-Da's Instagram [3], it can make humans and robots increasingly competitive in terms of creative processes. However, as technology develops, we as humans always try to learn along with the latest advances.

One of the fields that we can study in terms of creativity is Visual Communication Design. Visual Communication Design studies how to convey messages or information visually using design elements, such as images, typography, color, and layout. Visual Communication Design uses visual elements to convey information effectively, increase audience engagement and understanding. This is one of the design elements of images with digitization advances with its beauty and uniqueness [4]. Image elements are used to support promotions, visual identity, and convey information effectively [5], [6]. Images in graphic design help designers to convey visual communication in a more attractive display [7].

In the era of artificial intelligence, image elements with digitalization can be facilitated with Artificial Intelligence Image Generator. By having AI Image generators, it facilitates the designers in the designing process with attractive results, realistic adjusted to text input. Not only image elements, the overall content of the Visual Communication design can be generated effectively to promote the business. AI Image Generator can provide new forms of high-quality visual alternatives that can help the creative process of designers [8].

AI image generators can make it easier for users to create digital illustrations by simply typing a text description and letting the technology produce images based on the text description given. Text to image technology is also developing with its advantages in producing realistic images. The purpose of the research is to find out what applications are used in AI image generators, as well as to identify the results of digital illustrations using AI image generators. In accordance with this, researchers have been interested in conducting research on AI image generators for the past 5 years in order to determine how capable AI image generators are in creating digital illustrations.

2. Method

This study uses a secondary literature review method from the results of research conducted by previous researchers. The literature method in the form of a Systematic Literature Review (SLR) aims to collect existing evidence. Articles are compiled through a systematic review of the Google Scholar database, some are also assisted by Harzing's Publish or Perish software. Initial data collection amounted to 211 articles using the keywords AI, education, text to image, text to picture, AI Image Generation, Bing Image Creator, from research results spanning the period 2020 to 2024. The procedure used in the research is PRISMA (link). The final articles consisted of 25 articles for review by the authors.

3. Results and Discussion

3.1. Results

Articles found based on literature search using PRISMA. The relevant selection results are shown in Table 1.

Table 1. Relevant Paper Findings

Number	Year	Ref	Summary
	2021	[9]	GLIDE (Guided Language to Image Diffusion for Generation and Editing) uses a text-guided diffusion model to generate more photorealistic images. The model is equipped with editing capabilities to match the given description.
	2022	[10]	CLIPstyler is able to perform image style transfer using only a text description of the desired style, without requiring a style image reference by utilizing a pre-trained CLIP text image embedding model.
	2022	[11]	The quality of the images generated from the prompt should focus on the subject and style keywords; 2) It is recommended to generate 3-9 images with different seeds so that the results are more representative; 3) The number of literacy between 100-500 is enough to produce good images; 4) The AI generator can capture various painting styles well; 5) Real subjects produce better images.
	2022	[12]	CLIP (Contrastive Language-Image Pre-training) uses a two-stage approach consisting of a prior and a decoder model. The prior generates CLIP image embeddings from text captions, while the decoder generates images from those embeddings. The decoder uses a diffusion model technique to generate images in stages, CLIP as an image generator is capable of performing zero-shots that generate images for new concepts.
	2022	[13]	The Pathways Autoregressive Text-to-Image (Parti) model effectively generates high-quality photorealistic images from complex text descriptions.
	2022	[14]	The 4 P Model, namely: 1) Product: The Digital Image; 2) Person: The Practitioner; 3) Process: Iterative Prompt Engineering and Image Curation; 4) Press: The Emerging Text-to-Image Ecosystem. Rhodes' 4 P Model (product, person, process, and environment) to explore human creativity in text-based image generation, with iterative practices (refining, improving the product) and

Number	Year	Ref	Summary
			interactions of prompt engineering, as well as online communities that support creative practices.
	2022	[15]	A text-to-image diffusion model that combines the power of language models with diffusion models to produce images with a better degree of photorealism.
	2023	[16]	DreamBooth is capable of synthesizing images of subjects in various poses, and lighting conditions that are not present in the reference images, generating realistic images of subjects in new scenes, modifying the subject's properties from color or shape and generating artistic renderings.
	2023	[17]	Improved AI capabilities in generating images based on text according to the concept of objects, styles, characters with a variety of styles and poses and according to user desires.
	2023	[18]	A framework that does not require re-tuning, but leverages an internal representation of a text-guided diffusion model capable of translating sketches, rough drawings, and animations into realistic images, capable of modifying lighting and color quality.
	2023	[19]	ControlNet successfully adds conditional control to the text-to-image diffusion model efficiently, and can control Stable Diffusion with various inputs with or without text prompts. The results show that the ControlNet architecture can be applied to a wide range of conditions and facilitates relevant applications, more detailed control in image generation from text, which was previously difficult to achieve with conventional text-to-image diffusion models.
	2023	[20]	The workshop involving 15 participants revealed that the creative process inspired by AI encouraged educators to reflect on the unique nature of craft, AI to assist in ideation and visualization of text prompts that could enhance ideas.
	2023	[21]	Text from a word cloud containing artists, brands, or names used as inspiration in 72,980 Stable Diffusion prompts. Legal clarity on using Stable Diffusion training materials is critical for ethics in the Education environment, potential copyright infringement due to AI’s ability to reproduce artists’ work especially when users include the name of the artist’s work in the prompt.
	2023	[22]	Three AI Image Generators from Microsoft Bing Image Creator, Stable Diffusion, and Craiyon V3) have not been able to produce standard accuracy of human anatomy, namely skull or heart, and human brain images. The results still require a training dataset with more correct anatomical images for medical education and practice.
	2023	[23]	Exploration of 10 high school students with AI text-to-image generative models using Midjourney can enhance creativity and interaction between students and AI that emerge from dynamic relationships and in accordance with sociomaterial perspectives. Sociomaterial perspectives can be understood as dynamic learning and teaching processes of human and non-human groups.
	2023	[24]	AI-based image creation consists of three steps, namely: first, conceptual design is defined together with a set of keywords; second, the application of AI in image creation by designers efficiently explores the design and its visualization; third, utilizing image processing into a real project. The potential is applied to a historic urban center in Apulia, Southern Italy.
	2023	[25]	The capabilities of diffusion-based generative models, especially Stable Diffusion from Diffusion Explainers that converts text commands into high-resolution images. Diffusion Explainers is available online with video demonstrations to help users.
	2024	[26]	Building on existing finetuning-based personalization techniques, such as DreamBooth and TextualInversion, InstantBooth is not only adept at aligning

Number	Year	Ref	Summary
			language with images, preserving image quality, and maintaining image identity, but also increases the speed of image generation by 100x.
	2024	[27]	HyperDreamBooth is a hypernetwork capable of generating a small set of personalized weights from a single image of a person. The weights are used to control how the model converts text to images. HyperDreamBooth uses HyperNetwork to predict personalized weights from a single image, which are then used in a diffusion model to generate the corresponding image while preserving the details of the subject's identity, with high-quality results and a variety of styles.
	2024	[28]	T2I-Adapter achieves better control without requiring retraining of the base model, maintains high image quality, is able to adapt to diverse input conditions without sacrificing output quality, and can provide precise control over specific image attributes such as greater precision and flexibility to the user.
	2024	[29]	Students in the PPT group created more text and were interested in using Bing Chat to expand the text of the PPT slides. AI image descriptions from ChatGPT were used as prompts that effectively generated AI images from Bing Image Creator, students used prompts suggested by Bing Chat to refine the AI images.
	2024	[30]	AI-generated images show gender bias with underrepresentation of Black women in certain professions. AI-generated images associate certain occupations and genders with male or female dominance with generative AI models, such as DALL-E and Bing Image Creator.
	2024	[31]	Empathy in understanding the narratives of the experiences of individuals affected by crisis situations can facilitate AI algorithms in visualizing post-conflict urban landscapes using text to image models with DALL.E 2 and Bing Image Creator.
	2024	[32]	Bing Image Creator uses the DALL-E 2 model to depict a temperature change disaster expected to occur in the Tangier-Tetouan-Al Hoceima (TTA) region in 2099. The resulting image can be evaluated based on the social, economic, and natural characteristics of the region. Bing Image Creator can provide an alarming view and understand the impacts of economic and environmental losses due to rising temperatures in the region.
	2024	[33]	AI-generated images based on textual cues can be systematically analyzed using a social semiotic perspective and the notion of systemic functional linguistics adopting Kress and Van Leeuwen's grammar of visual design.

3.2. Discussion

Stable Diffusion is a modeling approach in fields such as: neurotechnology (technology designed to interact with the brain), text-to-image maker, and health image production (medical imaging). Stable Diffusion, specifically the Stable Diffusion XL (SDXL) base model, is advanced in generating high-quality images from textual descriptions by leveraging the principles of diffusion models and transformer architecture [34]. The Stable Diffusion method is enhanced with latent diffusion model (Latent Diffusion Model-LDM) for text-to-image synthesis. Image advancement with text-to-image improves the quality and control of the resulting image using composition inversion method of two modules, namely semantic inversion and spatial inversion [35]. Textual description with semantic inversion and spatial inversion, in this case, semantic inversion is concerned with understanding the changes in the meaning of words in processing natural language, while spatial inversion is used to preserve certain details in the resulting image.

GLIDE (Guided Language to Image Diffusion for Generation and Editing) is a diffusion model that can generate images from text descriptions based on the given text prompts, inpainting images from the process of editing images in certain parts and filling text prompts to fill deleted areas according to the surrounding style and lighting. GLIDE can use CLIP guidance to direct the diffusion process to ensure that the generated image matches the given text description [9].

CLIPstyler can make image style transfer with single text conditions with realistic texture in situations without image style reference [10]. The prompt technique in finding sentences from language models can produce the desired output from artists and other users in creating images from text. The success and failure modes of generating images from text prompts provide design guidelines to help users get better results from text to image generative [11].

CLIP (Contrastive Language-Image Pre-training) was developed by OpenAI to represent the combination of text and images by “zero-shot” generating images for new concepts [12]. Pathways Autoregressive Text-to-Image (Parti) model has several advantages compared to other text-to-image models such as DALL-E, CogView, diffusion-based models such as GLIDE and Imagen, namely Parti introduces a new holistic benchmark called PartiPrompts. Parti Prompts consists of more than 1600 English prompts to measure the model's ability in various categories and provide a more complete evaluation [13].

Creative expression of text-to-image art requires insight into the instructions used, the iterative process of rapid engineering, and the role of community resources. Thus, a holistic approach is needed in assessing creativity, which includes not only the final product but also the process and environment that create the art [14].

DrawBench, a benchmark for evaluating text-to-image models, shows that Image is preferred over other methods, such as VQ-GAN+CLIP, Latent Diffusion Models, GLIDE and DALL-E 2. Human raters prefer Imagen in terms of sample quality and image-text alignment and alignment [15].

DreamBooth is a text-to-image diffusion model that generates subject images using only a few reference images (usually 3-5) of a given subject, and produces results with a high degree of accuracy and specificity in a variety of contexts [16]. Custom Diffusion compared with DreamBooth and Textual Inversion methods with various datasets, adjusting single concepts, adjusting multiple concepts, text-image alignment and better visual similarity, although still difficult in complex compositions and inserting concepts simultaneously [17].

A new framework that leverages a pre-trained text-to-image diffusion model for text-guided text-to-image translation tasks. The results retain the meaningful layout of the input image while changing its appearance based on textual commands [18]. ControlNet is a neural network architecture that adds spatial conditioning control (the spatial aspect of an object) to text-to-image diffusion models, including pre-trained instruction-based image editing, thereby improving performance [19].

AI visualization of crafts only helps with ideas, as it is important to maintain the manual way of making crafts to provide a sensory and emotional experience, as well as ensuring future teachers are prepared to leverage AI while protecting the core value of craft education skills [20].

Text-to-image artificial intelligence has made a major breakthrough with the release of DALL-E and its open-source partner, Stable Diffusion, which can create visual artworks simply by providing a natural language description (prompt). Text-to-image AI has the potential to revolutionize the way art is taught, saving time and money on expression. But there are concerns about the ownership of works of art by artists [21].

AI Image Generator, although superior in speed, but the results for medical anatomy illustrations are less accurate. Therefore, the expertise of a human medical illustrator is needed to represent the nuances of anatomy [22]. Collaboration and critical reflection between teacher and student dialogue that creates an interactive space that encourages students to talk, explore, and help students understand the AI Image Generator [23].

Conceptual design defines the project vision and keywords, followed by Midjourney which produces design variations with an iterative process until the desired result is achieved. The final step is to transform the AI-generated images into CAD files of the 3D CAD design, then used to create a physical object through digital manufacturing techniques with 3D printing in the form of a cube pavilion for a city center in Apulia, Southern Italy that successfully integrates AI with a traditional architectural context [24].

Diffusion Explainer is a web-based interactive visualization tool that can explain how Stable Diffusion generates high-resolution images from prompt text. Users can discover the impact of prompt text on the image generation process with modern generative AI techniques [25]. InstantBooth can create high-quality personalized images, more consistently preserving subject identity, using learning techniques to extract identity features from reference images, producing faster and better images than other methods [26].

Writing weights into a diffusion model, coupled with fast refinement, HyperDreamBooth can generate a person's face in a variety of contexts and styles, while preserving high subject detail [27]. T2I-Adapter functions as an additional component that enriches the capabilities of the main text-to-image model, such as Stable Diffusion to be able to produce images [28].

Learning using generative AI, such as ChatGPT, Bing Chat to generate text, Bing Image Creator to generate images. Learning that can be applied from generative AI includes summarizing, creating concept maps, drawing, and asking questions. By learning generative, students can improve cognitive processes such as the ability to remember, understand, apply, analyze, evaluate, and create. In addition, it has been proven to improve cognitive, affective, and psychomotor aspects [29].

The image results in terms of dominant professions are shown as male gender, and there is a slight improvement in terms of showing sexualized female gender, as well as gender roles since childhood. DALL-E in the representation of Women in several professions such as doctors and CEOs. Bing Image Creator representation of Women in the professions of teachers, architects, engineers, and journalists who are depicted as younger, more attractive, and in more passive poses than men. Bing Image Creator of the representation of boys is associated with trucks or technical toys, while Women are associated with activities related to beauty and social norms [30].

AI-generated images can also bring images to life and help understand landscape transformation. Landscape narratives can preserve images and histories of cities among communities and in refugee camps affected by conflict [31]. Bing Image Creator uses OpenAI’s DALL-E 2 model trained on a dataset of images and text descriptions to generate images. DALL-E 2 consists of a Prior component that converts text to image representations, and a Decoder that converts the representations to realistic photos. DALL-E 2 uses CLIP’s to-image integration that generates the most appropriate captions for a given image [32]. The role of semiotics in the AI era to model the process of translating meaning between visual representations that are organized and aligned with textual descriptions [33].

3.3. A Digital Illustration an AI Image Generator

Diffusion-based methods in artificial intelligence are generative techniques used to create new, realistic data by gradually adding and removing noise. Diffusion-based methods work with a forward process, namely the original image is gradually added with noise until it becomes completely random data, aiming to gradually destroy the original data structure. Diffusion-based methods work with a backward process, namely the model is trained to gradually remove noise from the data that has been randomized, then return it to its original form or produce new, realistic data. This process involves learning parameters that make the model reverse the process of adding noise.

Image generation diffusion models with various techniques and variations developed to improve text-based quality and control have unique advantages and approaches. For example, GLIDE and Imagen focus on high image quality, while ControlNet and T2I-Adapter offer better control over the generation process. DreamBooth and TextualInversion create greater personalization, while CLIP provides a strong semantic understanding between text and images.

Diffusion-based methods have been used in image generation applications of DALL-E, Stable Diffusion, Midjourney, Adobe Firefly, Craiyon V3, and Bing Image Creator models that generate images from user-provided text descriptions, restore images by filling in missing or damaged parts of the image, enhance image resolution by adding previously missing details.

AI image generator uses artificial neural networks and deep learning techniques to understand and translate text input into corresponding visuals based on text descriptions. The quality of the generated images also depends heavily on the quality of the prompt or text description provided by the user.

Digital illustrations produced by AI image generators have excellent quality and realism and can be used for creative processes. The visual styles produced are very diverse, ranging from illustrations, portraits, cartoons by choosing the style that best suits the user's needs.

4. Conclusion

Artificial intelligence in digital illustration creates a very broad use of data and has the potential to misuse technology. As the impact, it can create misleading or harmful content to the society. Images generated by artificial intelligence sometimes use data from existing artwork without having permission, consequently, it can lead to copyright issues. Inconsistent images from artificial intelligence sometimes have errors and look strange when viewed up close which need further revisement. Therefore, it is very important to integrate Artificial Intelligence in digital illustration with the abilities of human artists so that they can be maximized into better digital illustration works. The projects should be combined between humans and artificial intelligence in order to produce a semiotic meaning of the real value of beauty.

References

[1] I. M. M. Yusa, Y. Yu, and T. Sovhyra, "Reflections on the Use of Artificial Intelligence in Works of Art," *Jadam*, vol. 2, no. 2, pp. 152–167, 2022, doi: 10.58982/jadam.v2i2.334.

[2] J. Chi, "The Evolutionary Impact of Artificial Intelligence on Contemporary Artistic Practices," *Commun. Humanit. Res.*, vol. 35, no. 1, pp. 6–11, 2024, doi: 10.54254/2753-7064/35/20240006.

[3] D. Ashton, "'People Don't Buy Art, They Buy Artists': Robot Artists – Work, Identity, and Expertise," *Converg. Int. J. Res. Into New Media Technol.*, vol. 30, no. 2, pp. 790–806, 2024, doi: 10.1177/13548565231220310.

[4] Y. Wang, "Illustration Art Based on Visual Communication in Digital Context," *Mob. Inf. Syst.*, vol. 2022, no. 1, p. 7364003, Jan. 2022, doi: https://doi.org/10.1155/2022/7364003.

[5] M. Mujahid, M. J. Hafiz, and A. M. Sari, "Media Desain Komunikasi Visual Sebagai Penunjang Promosi Pada SMK Yuppentek 7 Tangerang," *Mavib J.*, vol. 3, no. 1, pp. 43–52, 2022, doi: 10.33050/mavib.v3i1.1572.

[6] A. Bakar, "Komunikasi Visual Penjenamaan Dalam Upaya Membangun Citra Visual Identitas Baru Kota Palembang," *Besaung J. Seni Desain Dan Budaya*, vol. 8, no. 2, pp. 151–161, 2023, doi: 10.36982/jsdb.v8i2.2970.

[7] F. Mutaqin, "Penerapan Smart Communication Bot Dengan Model Chatgpt Dalam Proses Pemilihan Dan Penempatan Objek Dalam Desain Poster Komersil," *J. Dasarupa Desain Dan Seni Rupa*, vol. 5, no. 3, pp. 1–9, 2024, doi: 10.52005/dasarrupa.v5i3.169.

[8] L.-Y. Chiou, P.-K. Hung, R.-H. Liang, and C.-T. Wang, "Designing with AI: An Exploration of Co-Ideation with Image Generators," in *2023 ACM designing interactive systems conference*, 2023, pp. 1941–1954. doi: 10.1145/3563657.3596001.

[9] A. Nichol *et al.*, "Glide: Towards photorealistic image generation and editing with text-guided diffusion models," *arXiv Prepr. arXiv2112.10741*, 2021, doi: 10.48550/arXiv.2112.10741.

[10] G. Kwon and J. C. Ye, "Clipstyler: Image style transfer with a single text condition," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 18041–18050. doi: 10.1109/CVPR52688.2022.01753.

[11] V. Liu and L. B. Chilton, "Design guidelines for prompt engineering text-to-image generative models," in *2022 CHI conference on human factors in computing systems*, 2022, pp. 1–23.

[12] A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, and M. Chen, "Hierarchical Text-Conditional Image Generation with CLIP Latents," *arXiv Prepr. arXiv2204.06125*, 2022.

[13] J. Yu *et al.*, "Scaling autoregressive models for content-rich text-to-image generation," *arXiv Prepr. arXiv2206.10789*, vol. 2, no. 3, p. 5, 2022.

[14] J. Oppenlaender, "The creativity of text-to-image generation," in *25th international academic mindtrek conference*, 2022, pp. 192–202. doi: 10.1145/3569219.3569352.

[15] C. Saharia *et al.*, “Photorealistic text-to-image diffusion models with deep language understanding,” in *Advances in neural information processing systems*, 2022, vol. 35. [Online]. Available: https://api.semanticscholar.org/CorpusID:248986576

[16] N. Ruiz, Y. Li, V. Jampani, Y. Pritch, M. Rubinstein, and K. Aberman, “Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation,” in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 22500–22510. doi: 10.48550/arXiv.2208.12242.

[17] N. Kumari, B. Zhang, R. Zhang, E. Shechtman, and J.-Y. Zhu, “Multi-Concept Customization of Text-to-Image Diffusion,” in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 1931–1941. doi: 10.48550/arXiv.2212.04488.

[18] N. Tumanyan, M. Geyer, S. Bagon, and T. Dekel, “Plug-and-play diffusion features for text-driven image-to-image translation,” in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 1921–1930. doi: 10.48550/arXiv.2211.12572.

[19] L. Zhang, A. Rao, and M. Agrawala, “Adding conditional control to text-to-image diffusion models,” in *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023, pp. 3836–3847. doi: 10.48550/arXiv.2302.05543.

[20] H. Vartiainen and M. Tedre, “Using artificial intelligence in craft education: crafting with text-to-image generative models,” *Digit. Creat.*, vol. 34, no. 1, pp. 1–21, 2023.

[21] N. Dehouche and K. Dehouche, “What’s in a text-to-image prompt? The potential of stable diffusion in visual arts education,” *Heliyon*, vol. 9, no. 6, pp. 1–12, 2023, doi: https://doi.org/10.1016/j.heliyon.2023.e16757.

[22] G. P. J. C. Noel, “Evaluating AI-powered text-to-image generators for anatomical illustration: A comparative study,” *Anat. Sci. Educ.*, 2023.

[23] H. Vartiainen, M. Tedre, and I. Jormanainen, “Co-creating digital art with generative AI in K-9 education: Socio-material insights,” *Int. J. Educ. through art*, vol. 19, no. 3, pp. 405–423, 2023.

[24] G. Fallacara, M. P. Fanti, F. Parisi, N. Parisi, and V. Sangiorgio, “AI-driven image generation for enhancing design in digital fabrication: urban furnishings in historic city centres,” 2023.

[25] S. Lee *et al.*, “Diffusion explainer: Visual explanation for text-to-image stable diffusion,” *arXiv Prepr. arXiv2305.03509*, 2023, doi: 10.48550/arXiv.2305.03509.

[26] J. Shi, W. Xiong, Z. Lin, and H. J. Jung, “Instantbooth: Personalized text-to-image generation without test-time finetuning,” in *EEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 8543–8552.

[27] N. Ruiz *et al.*, “HyperDreamBooth: HyperNetworks for Fast Personalization of Text-to-Image Models,” in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul. 2024, pp. 6527–6536. doi: 10.48550/arXiv.2307.06949.

[28] C. Mou *et al.*, “T2i-adapter: Learning adapters to dig out more controllable ability for text-to-image diffusion models,” in *AAAI Conference on Artificial Intelligence*, Mar. 2024, vol. 38, no. 5, pp. 4296–4304. doi: 10.1609/aaai.v38i5.28226.

[29] M. Liu, L. J. Zhang, and C. Biebricher, “Investigating students’ cognitive processes in generative AI-assisted digital multimodal composing and traditional writing,” *Comput. Educ.*, vol. 211, p. 104977, 2024, doi: https://doi.org/10.1016/j.compedu.2023.104977.

[30] T. Sandoval-Martin and E. Martínez-Sanzo, “Perpetuation of Gender Bias in Visual Representation of Professions in the Generative AI Tools DALL· E and Bing Image Creator,” *Soc. Sci.*, vol. 13, no. 5, p. 250, 2024, doi: 10.3390/socsci13050250.

[31] G. Alkhateeb, J. Storie, and M. Külvik, “Post-Conflict Urban Landscape Storytelling: Two Approaches to Contemporary Virtual Visualisation of Oral Narratives,” *Land*, vol. 13, no. 4.

p. 406, 2024. doi: 10.3390/land13040406.

[32] A. Tahri, M. Beroho, I. Boulahfa, and K. Aboumaria, “AI for awareness: Involvement in climate change in the Tangier-Tetouan-Al Hoceima region (Morocco),” *E3S Web Conf.*, vol. 502, 2024, [Online]. Available: https://doi.org/10.1051/e3sconf/202450201007

[33] A. Ghazvineh, “An inter-semiotic analysis of ideational meaning in text-prompted AI-generated images,” *Lang. Semiot. Stud.*, vol. 10, no. 1, pp. 17–42, 2024, doi: doi:10.1515/lass-2023-0030.

[34] P. Avhad, “WordCanvas: Text-to-Image Generation,” *Int. J. Sci. Res. Eng. Manag.*, vol. 08, pp. 1–5, May 2024, doi: 10.55041/IJSREM32152.

[35] X. Zhang *et al.*, “Compositional Inversion for Stable Diffusion Models,” in *AAAI Conference on Artificial Intelligence*, Mar. 2024, vol. 38, no. 7 SE-AAAI Technical Track on Computer Vision VI, pp. 7350–7358. doi: 10.1609/aaai.v38i7.28565.