Diffusion Explainer: Visual Explanation for Text-to-image Stable Diffusion

Seongmin Lee* Benjamin Hoover *,† Hendrik Strobelt† Zijie J. Wang* ShengYun Peng* Austin Wright* Kevin Li* Haekyu Park* Haoyang Yang* Duen Horng (Polo) Chau*

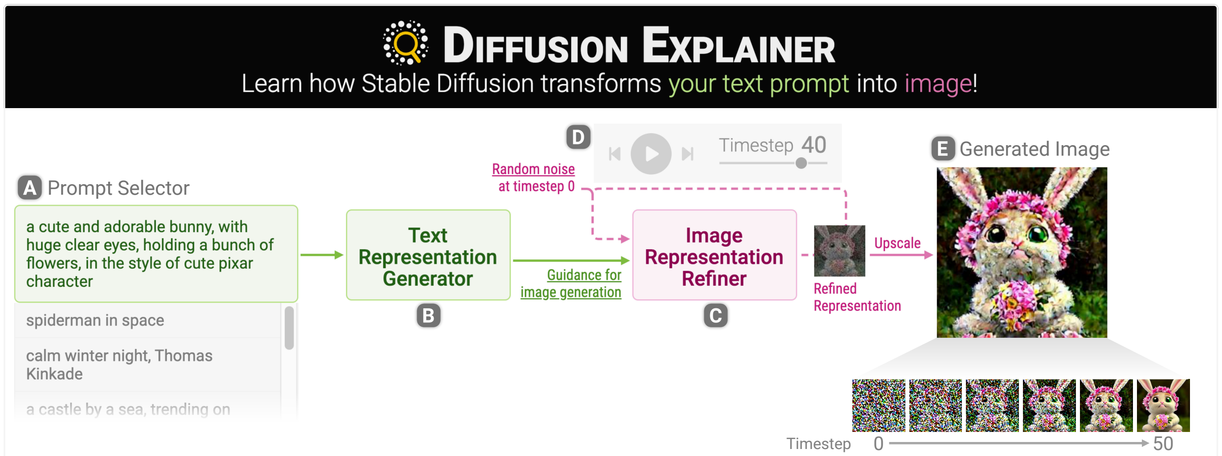

Figure 1: With Diffusion Explainer, users can visually examine how **(A)** *text prompt*, e.g., *"a cute and adorable bunny... pixar character"*, is encoded by **(B)** the *Text Representation Generator* into vectors to guide **(C)** the *Image Representation Refiner* to iteratively refine the vector representation of the image being generated. **(D)** The *Timestep Controller* enables users to review the the incremental improvements in image quality and adherence to the prompt over timesteps. **(E)** The final image representation is upscaled to a high-resolution image. Diffusion Explainer tightly integrates a visual overview of Stable Diffusion's complex components with detailed explanations of their underlying operations, enabling users to fluidly transition between multiple levels of abstraction through animations and interactive elements (see Figures 2 and 3).

Abstract

Diffusion-based generative models' impressive ability to create convincing images has captured global attention. However, their complex internal structures and operations often make them difficult for non-experts to understand. We present Diffusion Explainer, the first interactive visualization tool that explains how Stable Diffusion transforms text prompts into images. Diffusion Explainer tightly integrates a visual overview of Stable Diffusion's complex components with detailed explanations of their underlying operations, enabling users to fluidly transition between multiple levels of abstraction through animations and interactive elements. By comparing the evolutions of image representations guided by two related text prompts over refinement timesteps, users can discover the impact of prompts on image generation. Diffusion Explainer runs locally in users' web browsers without the need for installation or specialized hardware, broadening the public's education access to modern AI techniques. Our open-sourced tool is available at: https://poloclub.github.io/diffusion-explainer/. A video demo is available at https://youtu.be/Zg4gxdIWDds.

Index Terms: Human-centered computing—Visualization—Visualization systems and tools—Visualization toolkits;

1 Introduction

Diffusion-based generative models [31, 37, 43] like Stable Diffusion [43] and DALL-E [31] have captured global attention for their impressive image creation abilities, from AI developers, designers, to policymakers. The integration of Stable Diffusion in Lensa AI [35], a photo editing application that transforms selfies into different styles of artwork like anime and fantasy, led to a surge of 5.8 million downloads in the first week of December 2022 [13].

However, the popularity and progress of generative AI models have sparked ethical and social concerns [12, 14, 15, 44], such as accusations of artistic style theft by developers of AI image generators [14, 15]. Policymakers are also discussing ways to combat malicious data generation and revise copyright policies [1, 16, 17, 39]. There is an urgent need for individuals from many different fields to understand how generative AI models function and communicate effectively with AI researchers and developers [15, 20].

Key challenges in designing learning tools for Stable Diffusion. At the high level, Stable Diffusion iteratively refines *noise* into a high-resolution image's vector representation, guided by a text prompt. Internally, the prompt is tokenized and encoded into vector representations by the *Text Encoder* of the *CLIP* neural network [36]. With text representations' guidance, Stable Diffusion improves the image quality and adherence to the prompt by incrementally refining the image's vector representation using the *UNet* neural network [38] and the *Scheduler* algorithm [28] to predict and remove noise. The final image representation is upscaled to a high-resolution image [25]. The crux of learning about Stable Diffusion, therefore, originates from the complex interplay between the multiple neural network subcomponents, their intricate operations, and the iterative nature of image representation refinements. Such complex interactions are challenging even for experts to comprehend [47]. While some articles [4] and video lessons [5, 22] explain Stable Diffusion, they presume knowledge of machine learning and focus on model training and mathematical details.

Contributions. In this work, we contribute:

- **Diffusion Explainer, the first interactive visualization tool de-**

*Georgia Tech. {seongmin|bhoov|jayw|speng65|apwright|kevin.li|haekyu|alexanderyang|polo}@gatech.edu
†IBM Research. hendrik.strobelt@ibm.com

signed for non-experts to learn how Stable Diffusion transforms a text prompt into a high-resolution image (Fig. 1), overcoming key design challenges in developing interactive learning tools for Stable Diffusion (§ 3). Diffusion Explainer tightly integrates a visual overview of Stable Diffusion's complex structure with detailed explanations of their underlying operations (Fig. 2, Fig. 3) enabling users to fluidly transition between multiple abstraction levels through animations and interactive elements (§ 4.2).

- **Novel interactive visualization design** that enables users to discover the impacts of prompts on image generation. It compares how image representations evolve differently over refinement timesteps when guided by two related text prompts (Fig. 4), revealing how the keyword differences in prompts affect evolution trajectories that start from the same initial random noise. Since prompt engineering for Stable Diffusion has far been highly heuristic [26,32], Diffusion Explainer provides a new way for people to gain a better understanding of the impacts of text prompts on the complex image generation process (§ 4.3).
- **An open-sourced, web-based implementation** that broadens the public's education access to modern generative AI techniques without requiring any installation, advanced computational resources, or coding skills. We develop Diffusion Explainer as a web-based tool that runs locally in users' browsers, allowing large number of concurrent users to easily learn about Stable Diffusion directly on their laptops and tablets (§ 4.1). Diffusion Explainer is open-sourced[1] and available at the following public demo link: https://poloclub.github.io/diffusion-explainer/. A video demo is available at https://youtu.be/Zg4gxdIWDds.

2 Related Works

Interactive Visualizations for Explaining Deep Learning. Several web-based visualization tools, such as CNN Explainer [48], GAN Lab [24], and Adversarial-Playground [29], have been developed to help people understand deep learning. Google's Machine Learning Crash Course [19] employs Tensorflow Playground [41], which provides interactive visualizations for training simple neural networks. Moreover, various deep learning concepts are explained by many machine learning researchers and practitioners in their web articles [2, 18, 40] and blog posts [30] through the use of interactive visualizations. Inspired by the success of these previous works, we present Diffusion Explainer, an interactive visualization tool that explains text-to-image Stable Diffusion.

Explanations for Stable Diffusion. Online articles that explain Stable Diffusion [4, 7, 8, 21, 46] often assume that the audience has knowledge about machine learning and use jargon and mathematical equations that can be daunting for non-experts [4, 8, 21]. Tutorials in the form of Google Colab notebooks [33, 49] primarily focus on code implementation, while blog posts for beginners [7, 46] mostly address deployment and prompt engineering. To help users quickly understand how Stable Diffusion generates an image, Diffusion Explainer provides easy-to-understand explanations of its complex architecture and operations, integrating multiple abstraction levels through fluid animations and interactive elements.

3 Design Goals

By reviewing literature and online resources, we have identified four design goals (G1-G4) aimed at addressing the challenges people may face while learning about Stable Diffusion:

G1. Visual summary of Stable Diffusion. Stable Diffusion is comprised of multiple model components, each with a complex structure [37, 47]. Additionally, its incremental image generation, which refines noise into the vector representation of a high-resolution image, is a cyclic process that is uncommon in neural networks. Diffusion Explainer aims to provide an overview of the model architecture and data flow to help users quickly understand its overall structure (§ 4.2).

[1] https://github.com/poloclub/diffusion-explainer

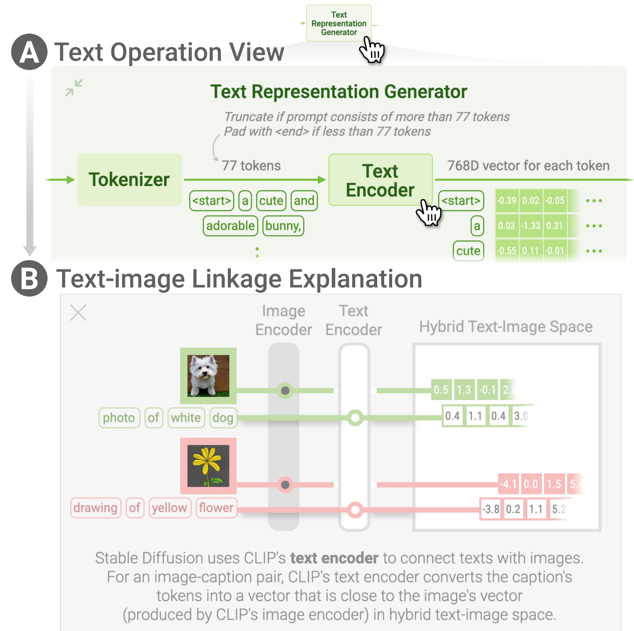

Figure 2: Diffusion Explainer tightly integrates different levels of abstractions to help users conceptually connect the overview of Stable Diffusion's structure with the underlying operations of each component. To learn how Stable Diffusion converts a text prompt into vector representations, users click the *Text Representation Generator*, which smoothly expands to (A) the *Text Operation View*, which explains how the prompt is split into tokens that are then encoded into vector representations. (B) The *Text-image Linkage Explanation* demonstrates how Stable Diffusion connects text and image, enabling text representations to guide the image generation process.

G2. Interactive interface tightly integrating different levels of abstraction. The image generation process of Stable Diffusion's image generation involves a complex interplay between multiple neural network subcomponents [36,37] (Fig. 2, Fig. 3), their intricate operations, and iterative image representation refinements. Such complex interactions are challenging even for experts to comprehend [47]. To effectively explain these low-level operations and help users conceptually connect them with a high-level overview, we design Diffusion Explainer to bridge multiple abstraction levels through fluid animations and interactive elements [24, 48] (§ 4.2.1, § 4.2.2).

G3. Visualizing how keywords in text prompts affect image generation. Stable Diffusion incrementally refines noise into the vector representation of a high-resolution image, while being guided by a text prompt. However, the refinement process, which involves multiple iterations of intricate vector computations, can be challenging to understand [37]. Due to the lack of understanding about how text prompts impact the refinements, writing prompts has been highly heuristic [26, 32]. We aim to visualize the refinement process for two text prompts that differ only in a few keywords to enable users to compare how image representations evolve differently when guided by each prompt. (§ 4.3).

G4. Broadening access via web-based deployment. As more and more individuals from different fields are now interested in understanding how generative AI models work [1, 15, 16, 39], we have developed Diffusion Explainer as a web-based tool that runs locally in users' web browsers without requiring any installation, specialized hardware, or coding skills [37]. This allows users to learn about this latest AI technology on their laptops or tablets (§ 4.1).

4 System Design and Implementation

4.1 Overview

Diffusion Explainer is an interactive visualization tool that explains how Stable Diffusion generates a high-resolution image from a text prompt, selected from the *Prompt Selector* (Fig. 1A). It incorporates an animation of random noise gradually refined and a *Timestep Controller* (Fig. 1D) that enables users to visit each refinement timestep. Diffusion Explainer consists of two views: *Architecture View* (§ 4.2) and *Refinement Comparison View* (§ 4.3). The Architecture View provides an overview of Stable Diffusion's architecture (G1), which can be expanded into details via user interactions (G2; Fig. 2, Fig. 3). The Refinement Comparison View visualizes the incremental image generation process for two related text prompts to allow users to discover how prompts affect image generation (G3). Diffusion Explainer is implemented using a standard web technology stack (HTML, CSS, JavaScript) and the D3.js [11] visualization library (G4). Diffusion Explainer has 13 text prompts based on the prompt template from *A Traveler's Guide to the Latent Space* [42]. Most prompts include popular keywords (e.g., *detailed*, *trending on artstation*) identified from literature and articles [9, 32, 34].

4.2 Architecture View

The Architecture View provides an overview (G1; Fig. 1) of how the *Text Representation Generator* (Fig. 1B) converts a text prompt into vector representations that guides the *Image Representation Refiner* (Fig. 1C) to incrementally refine noise into the vector representation of a high-resolution image. Clicking on the generators provides more details about their underlying operations (G2; Fig. 2, Fig. 3).

4.2.1 Text Representation Generator

The *Text Representation Generator* (Fig. 1B) converts text prompts into vector representations. Clicking on it expands to the *Text Operation View* (G2; Fig. 2A), that explains how the Tokenizer splits the prompt into tokens and how the Text Encoder encode the tokens into vector representations. Clicking on the Text Encoder displays the *Text-image Linkage Explanation* (G2; Fig. 2B), which visually explains how Stable Diffusion connects text and image by utilizing the CLIP [36] text encoder to generate text representations with image-related information.

4.2.2 Image Representation Refiner

The *Image Representation Refiner* (Fig. 1C) incrementally refines random noise into the vector representation of a high-resolution image that adheres to the input text prompt. Diffusion Explainer visualizes the image representation of each refinement step in two ways: (1) decoding it as a small image using linear operations [45] and (2) upscaling it to the Stable Diffusion's output resolution (Fig. 1E). Users expands the Image Representation Refiner to access the *Image Operation View* (G2; Fig. 3A), which explains how the UNet neural network [38] predicts the noise to be removed from the image representation to improve its adherence to the prompt. The predicted noise is weakened before removal.

The guidance scale hyperparameter, which controls the image's adherence strength to the text prompt, is described at the bottom, and further explained in the *Interactive Guidance Explanation* (G2; Fig. 3B) through a slider that allows users to experiment with different values, to better understand how higher values lead to stronger adherence of the generated image to the text prompt.

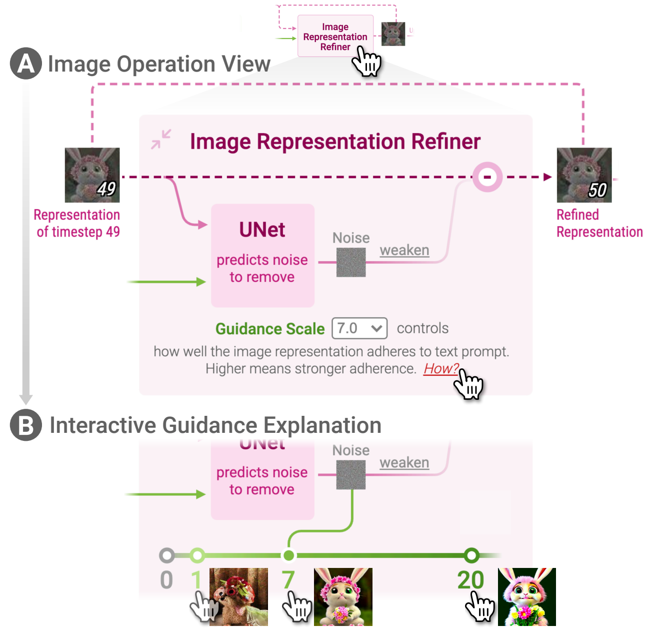

Figure 3: Users learn how Stable Diffusion incrementally refines noise into the vector representation of a high-resolution image that adheres to the text prompt by clicking the *Image Representation Refiner* in the high-level overview, which smoothly expands to (A) the *Image Operation View* that demonstrates how the noise is iteratively weakened and removed from the image representation as predicted by the UNet neural network. (B) The *Interactive Guidance Explanation* allows users to interactively experiment with different guidance scale values (0, 1, 7, 20) to better understand how higher values lead to stronger adherence of the generated image to the text prompt.

4.3 Refinement Comparison View

The *Refinement Comparison View* demonstrates how Stable Diffusion generates different images based on two related text prompts, helping users understand the impact of prompts on image generation (G3; Fig. 4). Each prompt in Diffusion Explainer is paired with a prompt that differs only in a few keywords (e.g., "a cute and adorable bunny..." vs. "a cute and adorable bunny... *pixar character*"). We use UMAP [27] to visualize the incremental refinement of image representations for each paired prompts, revealing how the keywords in prompts affect the evolution of image representations from the same initial random noise (G3).

5 Usage Scenarios

We present two usage scenarios for Diffusion Explainer, demonstrating how it may enhance user learning of Stable Diffusion. The scenarios highlight: (1) how practitioners can discover the impact of text prompts on image generation (§ 5.1); and (2) how non-experts can discern challenges in attributing AI-generated images (§ 5.2).

5.1 Discovering Prompts' Impact on Image Generation

Jenny is a graphic designer at a media company who wants to use generative AI models to create images in specific artistic styles, but she is uncertain how text prompts affect image generation. In particular, she wants to experiment with different styles while maintaining object composition consistency. Jenny activates the *Refinement Comparison View* (Fig. 4A), in Diffusion Explainer to compare two related text prompts and images generated from each. Both prompts begin with the phrase "a cute and adorable bunny", but only one includes "*in the style of cute pixar character*". The bunnies in both images have the same pose, but the *pixar* version is more cartoony

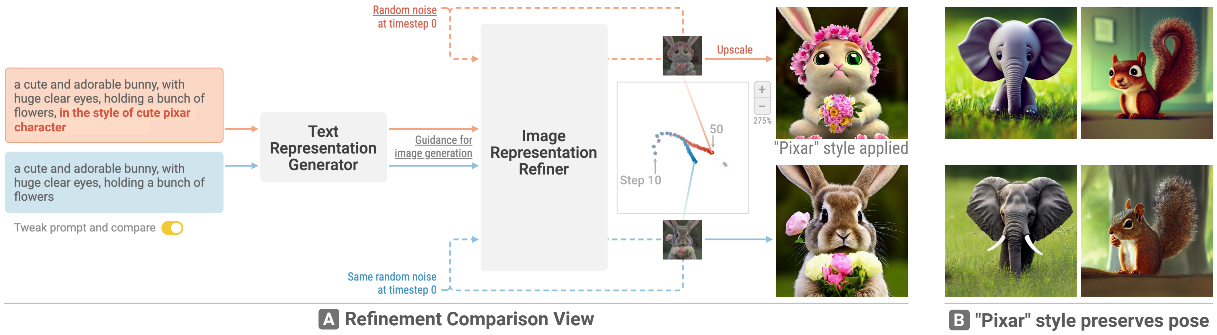

Figure 4: **(A)** The *Refinement Comparison View* enables users to discover the impacts of prompts on image generation by comparing how image representations evolve differently over refinement timesteps, using UMAP, when guided by two related text prompts. Adding *"pixar"* phrase changes the generated bunny's style to be more cartoony and vibrant in colors and textures while preserving its pose. **(B)** The same *pixar* phrase consistently preserves the poses of the elephant and squirrel.

and has more vibrant colors and textures, typical of characters in Pixar animations. Curious about whether the pose preservation is a coincidence, Jenny adds the same *pixar* phrase to prompts for an elephant and a squirrel (Fig. 4B) and notices that their poses are also preserved. Intrigued by the effect of the *pixar* phrase on image generation, she examines the trajectories of the image representations and discovers that adding the *pixar* phrase leads to only slight divergence.

Jenny wonders if other phrases may also similarly modify only styles while maintaining overall image compositions. To explore this, she asks her colleagues about commonly used "modifiers" keywords. Some suggest that repeating words such as "very very..." could produce better images by more reliably activating neural network regions associated with subject terms." [3, 32] Intrigued, Jenny compares the prompts "*a very very very very very beautiful cityscape*" [32] and "*a beautiful cityscape*." Surprisingly, the two prompts generate significantly different images. To understand why, Jenny analyzes the image representation trajectories and observes a detachment occurring at step 24, resulting in their final representations being much farther apart. From this, she concludes that the pose preservation of the *pixar* phrase is a unique characteristic attributable to its slight divergence and decides to identify more such keywords.

5.2 Discerning Challenges in Attributing AI Generations

Troy is a government policymaker responsible for creating policies related to AI-generated images in the entertainment and media industries. Recently, he has received numerous concerns from artists that their artwork has been exploited by AI models to create commercial products without their consent [6]. Troy is keen to help these artists be compensated for their contributions. In his research, he has learned about emergent tools that aim to help attribute AI-generated images to human artists [10, 23], which could potentially address artists' concerns. However, before proposing any policies, he needs to understand how and if such attribution may work.

Troy starts by launching Diffusion Explainer on his laptop, arriving at the Overview that describes how Stable Diffusion transforms a text prompt into a high-resolution image (Fig. 1). He realizes that the process of generating an image is iterative and involves refining noise into a vector representation of a high-resolution image that aligns with the text prompt. Curious about how the text prompt is processed, he clicks on the *Text Representation Generator* to expand it to the *Text Operation View* (Fig. 2A). Here, he discovers that the prompt is split into tokens and converted into vector representations. However, he is still unsure about how text guides image generation, so he displays the *Text-image Linkage Explanation* (Fig. 2B). Here, he learns that the text representations with image-related information act as a bridge between text and images.

Troy proceeds to explore the incremental refinement process of image representation by examining the *Image Operation View* (Fig. 3A). He discovers that each refinement step involves noise prediction and removal; UNet, a neural network, predicts the noise in the image representation of the step. He also learns about the *guidance scale*, a hyperparameter that adjusts how well the generated image adheres to the text prompt. Intrigued by the guidance scale, Troy accesses the *Interactive Guidance Explanation* (Fig. 3B). After experimenting with different guidance scale values, he observes that a guidance scale value of 7 generates a realistic image that closely follows the text prompt. In contrast, values of 1 and 20 result in images that are either difficult to interpret or overly exaggerated.

Troy has now gained a good understanding of the image generation process of Stable Diffusion, including the factors involved such as text prompts, guidance scale, and the link between text and image. Based on this understanding, he realizes that relying solely on image analysis, without considering text prompts, will be insufficient in determining how an artist's works have been used to create AI-generated images. Troy is of the opinion that more research is necessary to reliably identify attributions of AI-generated images.

6 Conclusion

We introduce Diffusion Explainer, the first interactive web-based visualization tool that explains how Stable Diffusion generates high-resolution images from text prompts. Our tool tightly integrates a visual overview of Stable Diffusion's complex components with detailed explanations of their underlying operations, enabling users to fluidly transition between multiple levels of abstraction through animations and interactive elements. Its novel interactive visualization design enables users to discover the impacts of prompts on image generation. Diffusion Explainer runs in modern web browsers and is open-sourced. We hope our work will inspire further research and development of visualization tools that helps enhance people's understanding of generative AI technologies so they may be used responsibly.

REFERENCES

[1] Copyright Office Launches New Artificial Intelligence Initiative. https://www.copyright.gov/newsnet/2023/1004.html, 2023. Accessed on: 2023-04-30.

[2] A. Agnihotri and N. Batra. Exploring Bayesian Optimization. *Distill*, 5(5):e26, 2020.

[3] Akshita. What are AI Image Generators? How Do They Work? https://narrato.io/blog/what-are-ai-image-generators-how-do-they-work/, 2023. Accessed on: 2023-04-30.

[4] J. Alammar. The illustrated Stable Diffusion. https://jalammar.github.io/illustrated-stable-diffusion/, 2022. Accessed on: 2023-04-30.

[5] J. Alammar. AI Art Explained: How AI Generates Images (Stable Diffusion, Midjourney, and DALLE). https://youtu.be/MXmacOUJUaw, 2023. Accessed on: 2023-04-30.

[6] AMELION. https://twitter.com/amelion_/status/1651193228677218304, 2023. Accessed on: 2023-04-26.

[7] Andrew. Absolute beginnerś guide to Stable Diffusion AI image. https://stable-diffusion-art.com/beginners-guide/, 2023. Accessed on: 2023-04-30.

[8] Andrew. How does Stable Diffusion work? https://stable-diffusion-art.com/how-stable-diffusion-work/, 2023. Accessed on: 2023-04-30.

[9] Andrew. Stable Diffusion prompt: a definitive guide. https://stable-diffusion-art.com/prompt-guide/, 2023. Accessed on: 2023-04-29.

[10] anton. Announcing Stable Attribution - A tool which lets anyone find the human creators behind AI generated images. https://twitter.com/atroyn/status/1622355473193381888, 2023. Accessed on: 2023-04-30.

[11] M. Bostock, V. Ogievetsky, and J. Heer. D^3 Data-driven Documents. *IEEE transactions on visualization and computer graphics*, 17(12):2301–2309, 2011.

[12] J. Brusseau. Acceleration AI Ethics, the Debate between Innovation and Safety, and Stability AI's Diffusion versus OpenAI's Dall-E. *arXiv preprint arXiv:2212.01834*, 2022.

[13] Ceci, L. Lensa AI Global App Downloads 2021-2022. https://www.statista.com/statistics/1350961/lensa-ai-app-downloads-worldwide/, 2023. Accessed on: 2023-04-29.

[14] L. Choudhary. Stable Diffusion is Now Accused of 'Stealing' Artwork. https://analyticsindiamag.com/stable-diffusion-is-now-accused-of-stealing-artwork/, 2022. Accessed on: 2023-04-30.

[15] P. Dixit. Meet The Three Artists Behind A Landmark Lawsuit Against AI Art Generators. https://www.buzzfeednews.com/article/pranavdixit/ai-art-generators-lawsuit-stable-diffusion-midjourney, 2023. Accessed on: 2023-04-30.

[16] A. Engler. Early thoughts on regulating generative AI like ChatGPT. *Brookings Institution*, 2023. Accessed on: 2023-04-30.

[17] A. G. Eshoo. Eshoo Urges NSA & OSTP to Address Unsafe AI Practices. https://eshoo.house.gov/media/press-releases/eshoo-urges-nsa-ostp-address-unsafe-ai-practices, 2022. Accessed on: 2023-04-30.

[18] G. Goh. Why Momentum Really Works. *Distill*, 2(4):e6, 2017.

[19] Google Developers. Machine Learning Crash Course with TensorFlow APIs. https://developers.google.com/machine-learning/crash-course/, 2022. Accessed on: 2023-04-30.

[20] J. Hendrix. Generative AI, Section 230 and Liability: Assessing the Questions. *Tech Policy Press*, 2023. Accessed on: 2023-04-30.

[21] Y. Hosni. Getting Started With Stable Diffusion. https://medium.com/towards-artificial-intelligence/getting-started-with-stable-diffusion-f343639e4931, 2022. Accessed on: 2023-04-30.

[22] J. Howard. From Deep Learning Foundations to Stable Diffusion. https://www.fast.ai/posts/part2-2023.html, 2023. Accessed on: 2023-04-30.

[23] J. Huber and A. Troynikov. Stable Attribution. https://www.stableattribution.com, 2023. Accessed on: 2023-04-30.

[24] M. Kahng, N. Thorat, D. H. Chau, F. Viégas, and M. Wattenberg. GAN Lab: Understanding Complex Deep Generative Models using Interactive Visual Experimentation. *IEEE Transactions on Visualization and Computer Graphics*, 25(1), 2019.

[25] D. P. Kingma and M. Welling. Auto-encoding Variational Bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[26] V. Liu and L. B. Chilton. Design Guidelines for Prompt Engineering Text-to-image Generative Models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pages 1–23, 2022.

[27] L. McInnes, J. Healy, and J. Melville. Umap: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv preprint arXiv:1802.03426*, 2018.

[28] A. Q. Nichol and P. Dhariwal. Improved Denoising Diffusion Probabilistic Models. In *International Conference on Machine Learning*, pages 8162–8171. PMLR, 2021.

[29] A. P. Norton and Y. Qi. Adversarial-Playground: A Visualization Suite Showing how Adversarial Examples Fool Deep Learning. In *Visualization for Cyber Security (VizSec), 2017 IEEE Symposium on*, pages 1–4. IEEE, 2017.

[30] C. Olah. colahś blog. http://colah.github.io, 2023. Accessed on: 2023-04-30.

[31] OpenAI. DALL-E 2. https://openai.com/product/dall-e-2, 2022. Accessed on: 2022-09-28.

[32] J. Oppenlaender. A Taxonomy of Prompt Modifiers for Text-to-Image Generation. *arXiv preprint arXiv:2204.13988*, 2022.

[33] S. Patil, P. Cuenca, N. Lambert, and P. v. Platen. Stable Diffusion with Diffusers. https://huggingface.co/blog/stable_diffusion, 2022. Accessed on: 2023-04-30.

[34] N. Pavlichenko and D. Ustalov. Best Prompts for Text-to-Image Models and How to Find Them. *arXiv preprint arXiv:2209.11711*, 2022.

[35] Prisma Labs. Lensa is an all-in-one image editing app that takes your photos to the next level. https://prisma-ai.com/lensa, 2023. Accessed on: 2023-04-29.

[36] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, et al. Learning Transferable Visual Models from Natural Language Supervision. In *International conference on machine learning*, pages 8748–8763. PMLR, 2021.

[37] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer. High-resolution Image Synthesis with Latent Diffusion Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10684–10695, 2022.

[38] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015.

[39] T. Ryan-Mosley. An early guide to policymaking on generative AI. *MIT Technology Review*, 2023. Accessed on: 2023-04-30.

[40] B. Sanchez-Lengeling, E. Reif, A. Pearce, and A. B. Wiltschko. A Gentle Introduction to Graph Neural Networks. *Distill*, 6(9):e33, 2021.

[41] D. Smilkov, S. Carter, D. Sculley, F. B. Viégas, and M. Wattenberg. Direct-manipulation Visualization of Deep Networks. *arXiv preprint arXiv:1708.03788*, 2017.

[42] E. Smith. A Traveler's Guide to the Latent Space. https://sweet-hall-e72.notion.site/A-Traveler-s-Guide-to-the-Latent-Space-85efba7e5e6a40e5bd3cae980f30235f#976ba690a0904431aac693d59830a92c, 2022. Accessed on: 2023-04-29.

[43] Stability AI. Stable Diffusion Public Release. https://stability.ai/blog/stable-diffusion-public-release, 2022. Accessed on: 2022-08-22.

[44] M. Sung. Lensa, the AI portrait app, has soared in popularity. But many artists question the ethics of AI art. *NBC News*, 2022. Accessed on: 2023-04-30.

[45] K. Turner. Decoding latents to RGB without upscaling. https://discuss.huggingface.co/t/decoding-latents-to-rgb-without-upscaling/23204/2, 2022. Accessed on: 2023-04-30.

[46] C. van den Bogaard. An introduction to Stable Diffusion. https://medium.com/sogetiblogsnl/an-introduction-to-stable-diffusion-efd5da6b3aeb, 2022. Accessed on: 2023-04-30.

[47] P. von Platen. Testing Stable Diffusion is hard. https://github.com/huggingface/diffusers/issues/937, 2022. Accessed on: 2023-04-30.

[48] Z. J. Wang, R. Turko, O. Shaikh, H. Park, N. Das, F. Hohman, M. Kahng, and D. H. Chau. CNN Explainer: Learning Convolutional Neural Networks with Interactive Visualization. *IEEE Conference on Visual Analytics Science and Technology (VAST)*, 2020.

[49] J. Whitaker. Grokking Stable Diffusion. https://colab.research.google.com/drive/1dlgggNa5Mz8sEAGU0wFCHhGLFooW_pf1?usp=sharing, 2022. Accessed on: 2023-04-30.