

# 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\*

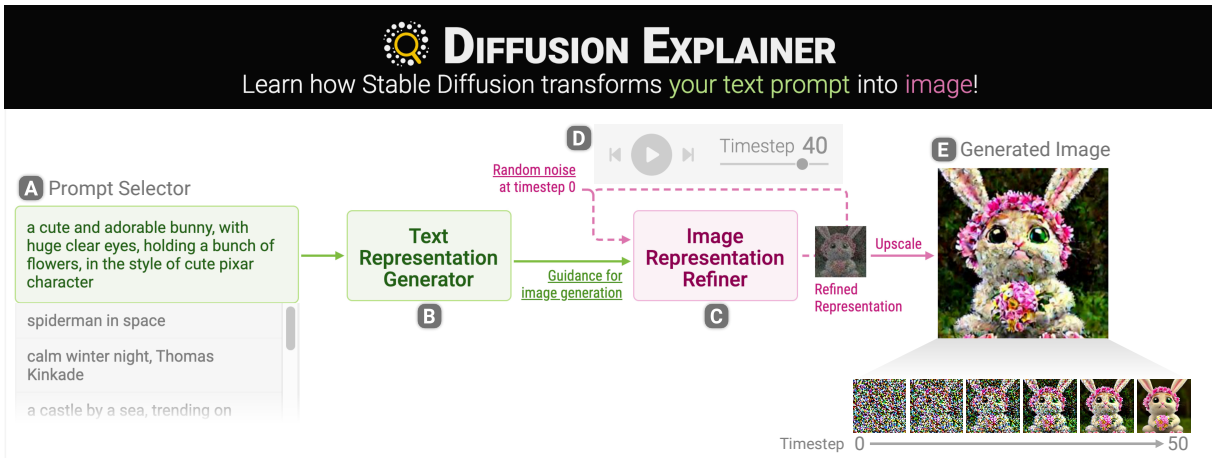


Figure 1: With Diffusion Explainer, users can visually examine how (A) text prompt, e.g., “a cute and adorable bunny... pixar character”, is encoded by (B) the Text Representation Generator into vectors to guide (C) the Image Representation Refiner to iteratively refine the vector representation of the image being generated. (D) The Timestep Controller enables users to review the 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 (Fig. 2, Fig. 3).

## ABSTRACT

Diffusion-based generative models’ impressive ability to create convincing images has garnered global attention. However, their complex structures and operations often pose challenges for non-experts to grasp. 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 structure with explanations of the underlying operations. By comparing image generation of prompt variants, users can discover the impact of keyword changes on image generation. A 56-participant user study demonstrates that Diffusion Explainer offers substantial learning benefits to non-experts. Our tool has been used by over 10,300 users from 124 countries at <https://poloclub.github.io/diffusion-explainer/>.

## 1 INTRODUCTION

Diffusion-based generative models [36, 43, 31] like Stable Diffusion [43] and DALL-E [31] have captured global attention for their impressive image creation abilities, from AI developers, designers, to policymakers. However, the popularity and progress of generative AI models have sparked social concerns [44, 9, 11, 12], such as accusations of artistic style theft by developers of AI image generators [11, 12]. Policymakers are also discussing ways to combat malicious data generation and revise copyright policies [14, 13, 38, 1].

\*Georgia Tech. {seongmin|bhoov|jayw|speng65|apwright|kevin.li|haekyu|alexanderyang|polo}@gatech.edu

†IBM Research. [hendrik.strobelt@ibm.com](mailto:hendrik.strobelt@ibm.com)

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 [12, 17].

### Key challenges in designing learning tools for Stable Diffusion.

At the high level, Stable Diffusion iteratively refines noise into a vector representation of a high-resolution image, guided by a text prompt. Internally, the prompt is tokenized and encoded into vector representations by CLIP Text Encoder [35]. With the text representations’ guidance, Stable Diffusion improves the image quality and adherence to the prompt by progressively denoising the image’s vector representation using the UNet neural network [37]. The final image representation is upscaled to a high-resolution image [24]. The crux of learning about Stable Diffusion stems from the complex interplay between multiple subcomponents, their intricate operations, and the iterative nature of refinements, which are challenging even for experts to grasp [47]. While some articles [3] and videos [19, 4] explain Stable Diffusion, they often assume notable knowledge of machine learning and focus on mathematical details.

In this work, we contribute:

- **Diffusion Explainer, the first interactive visualization tool designed for non-experts** to explain how Stable Diffusion transforms a text prompt into a high-resolution image, overcoming key design challenges in developing interactive learning tools for Stable Diffusion (Fig. 1). It tightly integrates a visual overview of Stable Diffusion’s complex structure with detailed explanations of their underlying operations via animations and interactive elements (Fig. 2, Fig. 3; §4.2). It also provides a new way to visualize the impact of keyword changes on the complex image generation process by comparing image generation of prompt variants (Fig. 4; §4.3).
- **Reflection and design lessons derived from human evaluation with 56 non-experts** showcase Diffusion Explainer’s substantial

advantages in explaining Stable Diffusion to non-experts, compared to the common blog post approach. The great majority preferred Diffusion Explainer, rating it easier to understand and more effective for improving their learning. We distill key design lessons for creating visualizations to educate non-experts on modern AI techniques (§ 5).

- **A web-based implementation** that broadens the public’s education access to modern generative AI techniques without requiring any installation, advanced hardware, or coding skills. Diffusion Explainer runs locally in users’ browsers, enabling a large number of concurrent users to learn directly on their devices (§ 4.1). Available at <https://poloclub.github.io/diffusion-explainer/>, Diffusion Explainer is open-source<sup>1</sup>. Having been used by over 10,300 users from 124 countries, Diffusion Explainer is making strides in democratizing AI education.

## 2 RELATED WORKS

**Interactive Visualizations for Explaining Deep Learning.** Early tools like ConvNetJS MNIST demo [23] and Tensorflow Playground [41] enabled users to experiment with simple models and datasets directly in their browsers. To explain more advanced techniques, researchers have developed interactive articles [16, 10, 27, 2, 39, 30], but they often assume prior machine learning knowledge. To address the needs of non-experts, interactive visualization tools such as CNN Explainer [48], GAN Lab [21], and Adversarial-Playground [29] were developed. Inspired by their success, we develop Diffusion Explainer as a web-based interactive visualization to broaden education access to Stable Diffusion.

**Explanations for Stable Diffusion.** Online articles that explain Stable Diffusion often assume machine learning expertise, using jargons and equations that can be daunting for non-experts [49, 3, 18, 6], while the articles for beginners [46, 5] mainly address deployment and prompt engineering. Additionally, they focus on either high-level structures [46, 5] or low-level operations [18, 6], overlooking the need for a comprehensive understanding [21, 48]. Google Colab tutorials [33, 50] require coding skills, posing challenges in learning. Diffusion Explainer enables easy experimentation without coding, offering clear explanations for Stable Diffusion’s architecture and operations through interactive elements.

## 3 DESIGN GOALS

By reviewing literature [36, 47], we established four design goals:

- G1. Visual summary of Stable Diffusion.** Stable Diffusion involves multiple complex model components [36, 47] and cyclic refinement from noise to the vector representation of a high-resolution image. Diffusion Explainer provides an overview of the model architecture and cyclic data flow to help users quickly understand its overall structure [20] (§ 4.2).
- G2. Interactive interface tightly integrating different abstraction levels.** Stable Diffusion’s image generation process is hard to comprehend due to a complex interplay between multiple subcomponents and their intricate underlying operations [47, 35, 36]. To effectively explain the low-level operations and conceptually connect them with a high-level overview, we bridge multiple abstraction levels through fluid animations and interactive elements [48, 21] (§ 4.2).
- G3. Visualizing how keyword changes in text prompts affect image generation.** Modifying a few keywords from prompts can unexpectedly lead to dramatic changes in generated images [25] (e.g., repeating “very” multiple times [32]), making it important for users to gain awareness and understanding of such impact on the generated images [15]. We visualize the refinement process for two text prompts that differ only in a

<sup>1</sup><https://github.com/poloclub/diffusion-explainer>

few keywords to compare how image representations evolve differently when guided by each prompt (§ 4.3).

- G4. Broadening access via web-based deployment.** As many individuals from various fields are interested in understanding generative AI [12, 13, 38], we develop Diffusion Explainer to run locally on users’ devices without requiring any installation, specialized hardware, or coding. To offer real-time interactive learning experiences [48], we pre-compute intensive processes on predetermined prompts (§ 4.1), instead of generating images for user-provided prompts using the nascent WebGPU technology [22], which is limited in both browser support and speed<sup>2</sup>. We open-source our code for easy extension to new prompts and hyperparameter settings.

## 4 SYSTEM DESIGN AND IMPLEMENTATION

### 4.1 Overview

Diffusion Explainer is a web-based interactive visualization that explains how Stable Diffusion generates a high-resolution image from a text prompt. It incorporates an animation of random noise gradually refined and a *Timestep Controller* (Fig. 1D) that enables users to visit each refinement timestep. From the *Prompt Selector* (Fig. 1A), users select one out of the 13 prompts that follow a template [42] and include popular keywords (e.g., *detailed*, *rending on arstition*) identified from literature [32, 7, 34]. Diffusion Explainer consists of two views: *Architecture View* (§ 4.2) tightly integrates a visual overview of Stable Diffusion’s architecture (G1) and the underlying operations via interactive elements and animation (G2), and *Refinement Comparison View* (§ 4.3) compares image generation of two related prompts to uncover the impact of prompt keywords on image generation (G3). Below the visualization, we provide text explanations for more details. Diffusion Explainer is implemented using a standard web technology stack (HTML, CSS, JavaScript) and the D3.js [8] library (G4).

### 4.2 Architecture View

The Architecture View shows a visual overview (G1, Fig. 1) of how *Text Representation Generator* (Fig. 1A) converts a text prompt into vector representations that guide *Image Representation Refiner* (Fig. 1B) to incrementally refine noise into the vector representation of a high-resolution image. Users can click the generators to expand them into more details about their underlying operations.

The *Text Representation Generator* converts a text prompt into vector representations. Users can click and expand it to the *Text Operation View* (G2, Fig. 2A) to learn that the prompt is split into tokens, and then the tokens are encoded into vector representations using Text Encoder. Clicking on the Text Encoder displays the *Text-image Linkage Explanation* (G2; Fig. 2B), which illustrates that CLIP [35] text encoder generates text representations with image-related information, which is crucial for guiding image generation.

The *Image Representation Refiner* incrementally refines random noise into the vector representation of an image that adheres to the text prompt. The image representation of each refinement step is visualized by (1) decoding into a small image using linear operations [45] and (2) upscaling to Stable Diffusion’s output resolution. Clicking the Image Representation Refiner expands it to the *Image Operation View* (G2; Fig. 3A), which explains that the refinement consists of noise prediction and 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). It provides a

<sup>2</sup>At the time of writing, WebGPU is only supported in Chrome browser and requires restricted settings to run Stable Diffusion [26]. Generating a single image takes at least a few minutes, with even longer times on CPUs. Training Stable Diffusion, which is even more computationally demanding, is therefore infeasible.

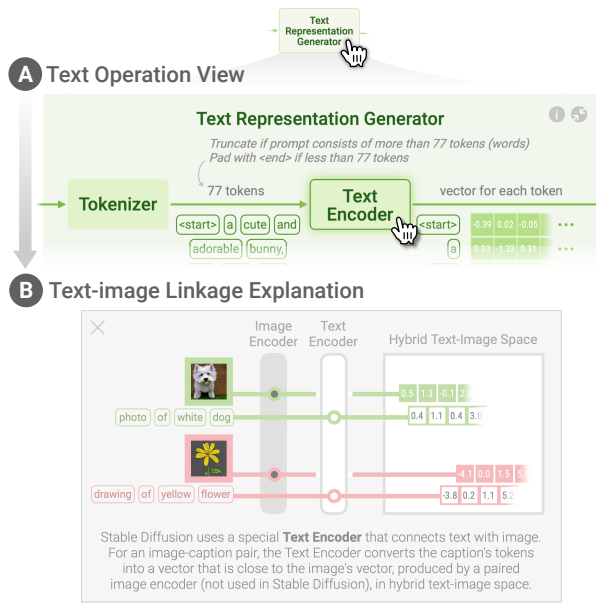


Figure 2: To learn how Stable Diffusion converts a text prompt into vector representations, users click the *Text Representation Generator*, which smoothly expands to (A) the *Text Operation View*, which explains how the prompt is split into tokens and encoded into vector representations. (B) The *Text-image Linkage Explanation* demonstrates how Stable Diffusion bridges text and image, enabling text representations to guide the image generation process.

slider to experiment with different guidance scale values to better understand how higher values lead to stronger adherence.

### 4.3 Refinement Comparison View

The *Refinement Comparison View* demonstrates how modifying a few keywords in prompt can significantly change the generated images (G3; Fig. 4). Each prompt in Diffusion Explainer is paired with a prompt that differs only in a few keywords; for example, an original prompt “a cute and adorable bunny... **pixar character**” is paired with the prompt variant “a cute and adorable bunny...” (keywords highlighted in **bold**). We smoothly animate the transitions between the Refinement Comparison View and the Architecture View, preserving the location of the original prompt and the model components, while fading in the prompt variant from underneath the original prompt.

We visualize the *Evolution Trajectory* of image representations for the paired prompts to reveal how prompt keywords affect the evolution of image representations (G3). Each point on the Evolution Trajectory corresponds to an image representation at each timestep, and their progression over timesteps is animated. We compute 2-dimensional representations of the image representations across all timesteps, text prompts, guidance scales, and random seeds using UMAP<sup>3</sup> [28].

## 5 HUMAN EVALUATION

To evaluate the effectiveness of Diffusion Explainer, we conducted a user study. We followed a within-subjects design [40] to compare Diffusion Explainer with other publicly accessible explanations.

<sup>3</sup>For UMAP hyperparameters, we conducted extensive testing over a wide range of `n_neighbors` (5 to 30), `min_dist` (0.1 to 0.99), and random seed (0, 1, 2), and did not observe a significant impact on the local structure. Hence, we employed a fixed configuration of `n_neighbors` of 15, `min_dist` of 0.99, and random seed of 0.

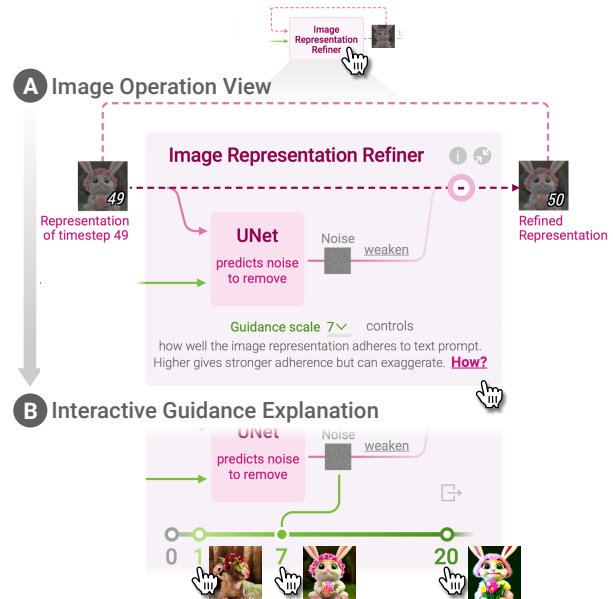


Figure 3: Users learn how Stable Diffusion refines noise into a high-resolution image’s vector representation aligned with the text prompt by clicking the *Image Representation Refiner* to smoothly expand to (A) the *Image Operation View* that demonstrates how noise is predicted and removed from the image representation. (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.

### 5.1 Procedure

We recruited participants from Prolific<sup>4</sup>, an online recruiting platform. After participants signed a consent document, we conducted a background survey. To include only non-experts interested in image generative AI, we asked participants to indicate their interest in image generative AI and to self-report their knowledge of AI and image generative AI on a scale from 1 to 5. We qualified only those who expressed interest and self-identified as having little knowledge of AI and generative AI (rated as 1: *Don’t know what it is*, 2: *Heard of it only*, or 3: *occasional users of AI-powered tools*).

Each participant then used Diffusion Explainer and a top-ranked blog post<sup>5</sup> on Google, to learn about Stable Diffusion. We excluded other sources (e.g., videos, Google Colab tutorials) as they were promotional, not free, or required advanced hardware and coding skills. To counterbalance the order effect, half of the participants began with Diffusion Explainer, while the other half used it after the blog post. To verify tool engagement, participants answered three simple quiz questions after using each tool<sup>6</sup>. We considered only responses from participants who answered at least two questions correctly for both tools. After using the two tools, participants rated them by answering 5-point Likert-scale questions.

The study lasted approximately 40 minutes per participant, with each compensated \$10.00. Fifty-six participants passed the knowledge and engagement screening. On average, their self-evaluated knowledge level in AI was 2.79, and in image generative AI was 2.41. Participants came from diverse fields, including art, education, public administration, health care, real estate, finance, retail, construction, and manufacturing.

<sup>4</sup><https://www.prolific.com>

<sup>5</sup><http://jalamar.github.io/illustrated-stable-diffusion/>

<sup>6</sup>Quizzes are easy for participants who used the tools, e.g., “Select all guidance scale values supported by Diffusion Explainer”

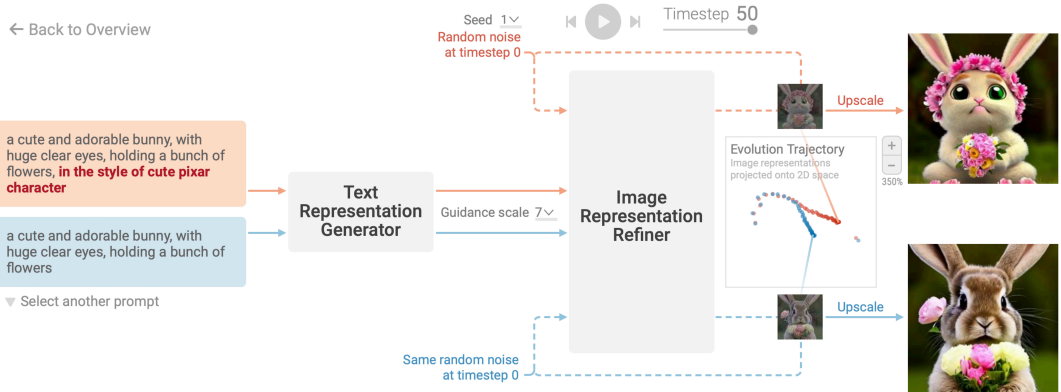


Figure 4: The *Refinement Comparison View* enables users to discover the impacts of prompts on image generation by comparing how image representations evolve differently over refinement timesteps, using UMAP, when guided by two related text prompts. Adding “*pixar*” phrase changes the generated bunny’s style to be more cartoony and vibrant in colors and textures while preserving its pose.

5.2 Results and Design Lessons

Fig. 5 summarizes the participants’ responses comparing the usability of the two tools. Overall, the great majority preferred Diffusion Explainer over the blog post (44 out of 56 participants). They found Diffusion Explainer more enjoyable, more helpful for improving their understanding, and easier to understand, even without expertise in AI.

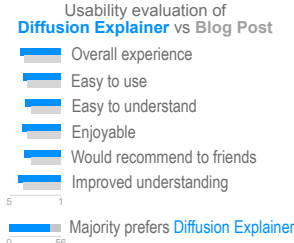


Figure 5: Diffusion Explainer more usable than blog post.

**Non-experts easily used Diffusion Explainer, improving understanding of Stable Diffusion.** We are excited to observe the high ratings given to all of Diffusion Explainer’s features (Fig. 6), showcasing its substantial educational advantages over the blog post (Fig. 5, Fig. 7). For example, when explaining how text prompts guide image generation, Diffusion Explainer received an average score of 4.18, notably surpassing the blog post’s 3.55 (Fig. 7).

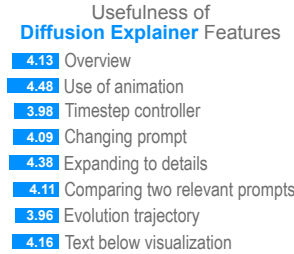


Figure 6: All features were rated highly.

**Thoughtful design crucial for bridging abstraction levels (G1, G2).** Our study confirms the key benefit of interactive design over a static post that can overwhelm users by presenting all information at once. Participants appreciated Diffusion Explainer’s high-level overview (4.13) and animated expansions into details (4.48). Our decision not to delve into advanced concepts, such as UNet’s architecture and the training of Stable Diffusion, as explained in depth in the blog post (Fig. 7), maintains Diffusion Explainer’s accessibility to non-experts. One participant noted, “*layers of the UNet noise predictor and jargons made [the blog post] too overwhelming and technical.*” This underscores the importance of carefully selecting details to include in visualizations to balance the amount of information with the users’ expertise.

**Interactive visualization offers unique learning benefits for understanding hyperparameters (G2).** Participants found Diffusion Explainer substantially more effective than the blog post in explaining the guidance scale (4.29 vs. 2.91) and random seed (4.25 vs. 2.88) (Fig. 7). They appreciated the ability to rapidly experiment with Diffusion Explainer. One participant noted, “[*it*] is like a

*cool game changing settings to get good-looking pictures for each prompt, helping me learn about making images from text.*” Also, participants could grasp the intricate interaction between prompts and hyperparameters that cannot be easily replicated by a static blog post. Another participant mentioned, “*The guidance scale for the best image quality depends on the prompt. I should decide the value carefully for each prompt.*”

Comparing prompt variants helps non-experts understand impact of prompt keywords (G3).

As a first tool designed to help non-experts learn about how slight changes in a prompt can significantly influence image generation, we are thrilled to learn that Diffusion Explainer’s Refinement Comparison View was favorably rated. In particular, prompt comparison received an average score of 4.11, and the Evolution Trajectory powered by UMAP [28], a technique often considered for advanced analysis, was also well-rated (3.96). The effectiveness is further demonstrated by Diffusion Explainer’s substantially higher rating (4.16 vs. 2.86) for improving understanding of the impact of prompt modification on image generation (Fig. 7). A participant remarked, “*I like how Diffusion Explainer lets me see how slight changes in prompts affect the style and quality of pictures. It’s cool to compare two different prompts and see how images and representations develop over time.*”

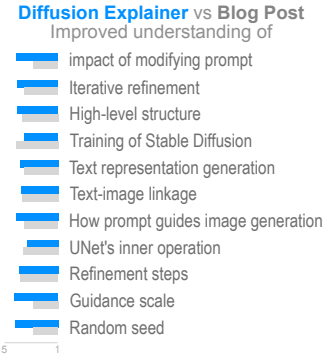


Figure 7: Diffusion Explainer more effective than blog post for learning most Stable Diffusion concepts.

6 CONCLUSION

We have introduced Diffusion Explainer, the first interactive web-based visualization tool that explains how Stable Diffusion generates high-resolution images from text prompts. Our tool seamlessly integrates a visual overview with detailed explanations of the underlying operations through animations and interactive elements. Its innovative design uncovers the impacts of prompt keywords on image generation. A user study with 56 non-experts demonstrates the superiority of Diffusion Explainer over a popular blog post. We hope our research inspires further development of visualizations to enhance people’s understanding of modern AI technologies.

## REFERENCES

- [1] Copyright Office Launches New Artificial Intelligence Initiative. <https://www.copyright.gov/newsnet/2023/1004.html>, 2023. Accessed on: 2023-04-30. 1
- [2] A. Agnihotri and N. Batra. Exploring Bayesian Optimization. *Distill*, 5(5):e26, 2020. 2
- [3] J. Alammar. The illustrated Stable Diffusion. <https://jalammar.github.io/illustrated-stable-diffusion/>, 2022. Accessed on: 2023-04-30. 1, 2
- [4] J. Alammar. AI Art Explained: How AI Generates Images (Stable Diffusion, Midjourney, and DALLE). <https://youtu.be/MXmacOUJUaw>, 2023. Accessed on: 2023-04-30. 1
- [5] Andrew. Absolute beginners guide to Stable Diffusion AI image. <https://stable-diffusion-art.com/beginners-guide/>, 2023. Accessed on: 2023-04-30. 2
- [6] Andrew. How does Stable Diffusion work? <https://stable-diffusion-art.com/how-stable-diffusion-work/>, 2023. Accessed on: 2023-04-30. 2
- [7] Andrew. Stable Diffusion prompt: a definitive guide. <https://stable-diffusion-art.com/prompt-guide/>, 2023. Accessed on: 2023-04-29. 2
- [8] M. Bostock, V. Ogievetsky, and J. Heer. D<sup>3</sup> Data-driven Documents. *IEEE transactions on visualization and computer graphics*, 17(12):2301–2309, 2011. 2
- [9] 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. 1
- [10] S. Carter and M. Nielsen. Using artificial intelligence to augment human intelligence. *Distill*, 2017. <https://distill.pub/2017/aia>. doi: 10.23915/distill.00009 2
- [11] 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. 1
- [12] 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. 1, 2
- [13] A. Engler. Early thoughts on regulating generative AI like ChatGPT. *Brookings Institution*, 2023. Accessed on: 2023-04-30. 1, 2
- [14] 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. 1
- [15] Y. Feng, X. Wang, K. K. Wong, S. Wang, Y. Lu, M. Zhu, B. Wang, and W. Chen. Promptmagician: Interactive prompt engineering for text-to-image creation. *IEEE Transactions on Visualization and Computer Graphics*, 2023. 2
- [16] G. Goh. Why Momentum Really Works. *Distill*, 2(4):e6, 2017. 2
- [17] J. Hendrix. Generative AI, Section 230 and Liability: Assessing the Questions. *Tech Policy Press*, 2023. Accessed on: 2023-04-30. 1
- [18] 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. 2
- [19] J. Howard. From Deep Learning Foundations to Stable Diffusion. <https://www.fast.ai/posts/part2-2023.html>, 2023. Accessed on: 2023-04-30. 1
- [20] M. Kahng, P. Y. Andrews, A. Kalro, and D. H. Chau. Activis: Visual exploration of industry-scale deep neural network models. *IEEE TVCG*, 24(1):88–97, 2017. 2
- [21] 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. 2
- [22] J. B. Kai Ninomiya, Brandon Jones. *Webgpu*. <https://www.w3.org/TR/webgpu/>, 2024. 2
- [23] A. Karpathy. ConvNetJS: Deep Learning in your browser. <https://cs.stanford.edu/people/karpathy/convnetjs/index.html>, 2014. Accessed on: 2023-11-26. 2
- [24] D. P. Kingma and M. Welling. Auto-encoding Variational Bayes. *arXiv preprint arXiv:1312.6114*, 2013. 1
- [25] V. Liu and L. B. Chilton. Design Guidelines for Prompt Engineering Text-to-image Generative Models. In *CHI*, pp. 1–23, 2022. 2
- [26] Machine Learning Compilation. Web Stable Diffusion. <https://mlc.ai/web-stable-diffusion>, 2024. Accessed on: 2024-04-28. 2
- [27] A. Madsen. Visualizing memorization in rnns. *Distill*, 2019. <https://distill.pub/2019/memorization-in-rnns>. doi: 10.23915/distill.00016 2
- [28] L. McInnes, J. Healy, and J. Melville. Umap: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv preprint arXiv:1802.03426*, 2018. 3, 4
- [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, pp. 1–4. IEEE, 2017. 2
- [30] C. Olah. colah’s blog. <http://colah.github.io>, 2023. Accessed on: 2023-04-30. 2
- [31] OpenAI. DALL-E 2. <https://openai.com/product/dall-e-2>, 2022. Accessed on: 2022-09-28. 1
- [32] J. Oppenlaender. A Taxonomy of Prompt Modifiers for Text-to-Image Generation. *arXiv preprint arXiv:2204.13988*, 2022. 2
- [33] S. Patil, P. Cuenca, N. Lambert, and P. v. Platen. Stable Diffusion with Diffusers. [https://huggingface.co/blog/stable\\_diffusion](https://huggingface.co/blog/stable_diffusion), 2022. Accessed on: 2023-04-30. 2
- [34] N. Pavlichenko and D. Ustalov. Best Prompts for Text-to-Image Models and How to Find Them. *arXiv preprint arXiv:2209.11711*, 2022. 2
- [35] 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*, pp. 8748–8763. PMLR, 2021. 1, 2
- [36] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer. High-resolution Image Synthesis with Latent Diffusion Models. In *CVPR*, pp. 10684–10695, 2022. 1, 2
- [37] 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*, pp. 234–241. Springer, 2015. 1
- [38] T. Ryan-Mosley. An early guide to policymaking on generative AI. *MIT Technology Review*, 2023. Accessed on: 2023-04-30. 1, 2
- [39] B. Sanchez-Lengeling, E. Reif, A. Pearce, and A. B. Wiltschko. A Gentle Introduction to Graph Neural Networks. *Distill*, 6(9):e33, 2021. 2
- [40] H. J. Seltman. *Experimental design and analysis*, 2012. 3
- [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. 2
- [42] E. Smith. A Traveler’s Guide to the Latent Space, 2022. Accessed on: 2023-04-29. 2
- [43] Stability AI. Stable Diffusion Public Release. <https://stability.ai/blog/stable-diffusion-public-release>, 2022. Accessed on: 2022-08-22. 1
- [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. 1
- [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. 2
- [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. 2
- [47] P. von Platen. Testing Stable Diffusion is hard. <https://github.com/huggingface/diffusers/issues/937>, 2022. Accessed on: 2023-04-30. 1, 2
- [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. 2
- [49] L. Weng. What are diffusion models? [lilianweng.github.io](https://lilianweng.github.io), Jul 2021. 2
- [50] J. Whitaker. Grokking Stable Diffusion. [https://colab.research.google.com/drive/1dgggNa5Mz8sEAGU0wFCHhGLF0w\\_pfl?usp=sharing](https://colab.research.google.com/drive/1dgggNa5Mz8sEAGU0wFCHhGLF0w_pfl?usp=sharing), 2022. Accessed on: 2023-04-30. 2