# Memory Triggers: Unveiling Memorization in Text-To-Image Generative Models through Word-Level Duplication

**Ali Naseh,**[1] **Jaechul Roh,**[1] **Amir Houmansadr**[1]

[1] University of Massachusetts Amherst
anaseh@cs.umass.edu, jroh@umass.edu, amir@cs.umass.edu

## Abstract

Diffusion-based models, such as the Stable Diffusion model, have revolutionized text-to-image synthesis with their ability to produce high-quality, high-resolution images. These advancements have prompted significant progress in image generation and editing tasks. However, these models also raise concerns due to their tendency to memorize and potentially replicate exact training samples, posing privacy risks and enabling adversarial attacks. Duplication in training datasets is recognized as a major factor contributing to memorization, and various forms of memorization have been studied so far. This paper focuses on two distinct and underexplored types of duplication that lead to replication during inference in diffusion-based models, particularly in the Stable Diffusion model. We delve into these lesser-studied duplication phenomena and their implications through two case studies, aiming to contribute to the safer and more responsible use of generative models in various applications.

## Introduction

*Diffusion-based* models (Sohl-Dickstein et al. 2015) have demonstrated outstanding ability in producing high-quality images, both *unconditionally* (Ho, Jain, and Abbeel 2020) and *conditionally* (Rombach et al. 2022; Nichol et al. 2021). The *Stable Diffusion* model (Rombach et al. 2022), a type of conditional diffusion model, alongside other generative models like DALLE-3 (Shi et al. 2020) and Midjourney (Midjourney 2022), has significantly advanced the field of text-to-image synthesis. These models excel in creating high-resolution images (Ramesh et al. 2022) and in image editing (Kim and Ye 2021).

Memorization across various machine learning models has been extensively researched (Carlini et al. 2019; Zhang et al. 2021). Such memorization can pose privacy risks, potentially enabling attacks such as membership inference (Shokri et al. 2017) or data extraction (Carlini et al. 2021). Despite the capabilities of diffusion-based models, including the Stable Diffusion model, to produce high-quality images, they occasionally exhibit tendencies to memorize and replicate exact training samples or significant portions thereof (Carlini et al. 2023; Somepalli et al. 2023a,b). Somepalli et al. (2023b) suggest that text conditioning contributes more to memorization than image-only contexts. Previous research has shown that training sample duplication could be a significant cause of such replication during inference.

This paper delves into two particular types of text-conditioned training sample duplication: the first involves duplication where the images, along with their corresponding texts containing specific keywords, are repeated; the second type pertains to duplication of image-text pairs where the images contain specific objects and the texts include specific keywords. We posit that this nuanced form of duplication may heighten the models' vulnerability to various attacks. With the burgeoning popularity of text-to-image generative models, a detailed scrutiny of their memorization propensities becomes increasingly crucial. We explore these two types of duplication by analyzing two case studies that shed light on their dynamics and implications.

## Related Works

### Memorization in Large Language Models

In the domain of Large Language Models (LLMs), there is an escalating challenge pertaining to the inadvertent disclosure of confidential information leading to model memorization (Lee et al. 2023; Carlini et al. 2021, 2023; Biderman et al. 2023). Carlini et al. (2022) perform in-depth analyses using quantitative methods, whereas Carlini et al. (2021) conduct their studies utilizing qualitative techniques. One of the pivotal factor causing this scenario is the replication inherent within training datasets, potentially causing the language model to generate text that mirrors already existing content (Lee et al. 2023). A recent contribution by Biderman et al. (2023) posit that such memorization manifests due to the average of the training dataset, while Biderman et al. (2023) illustrate its occurrence at specific training data point.

### Memorization in Diffusion Models

Recent studies (Carlini et al. 2023; Somepalli et al. 2023b,a) demonstrate techniques to create alike or almost identical images using both conditional and unconditional diffusion models. Specifically, Somepalli et al. (2023a) highlight that diffusion models can produce images with objects similar to those in the training data, in which the process is termed as 'replication' (Somepalli et al. 2023a). In parallel, Carlini et al. (2023) demonstrate the model's proficiency in retriev-

ing near-identical images from the training set by analyzing clusters of generated samples. Recent study by Somepalli et al. (2023b) posit that while data replication stemming from duplication might be infrequent in unconditional diffusion models, textual conditioning can notably augment the likelihood of model memorization. Echoing the occurrence from prior research that data duplication underlies such memorization, Webster et al. (2023) introduce an algorithmic approach to detect such repetitions.

# Background

## Diffusion Model

In the context of deep generative models, Denoising Diffusion Probabilistic Models (Ho, Jain, and Abbeel 2020), commonly referred to as unconditional diffusion models, iteratively engage in the act of noise addition (forward process) and subsequent noise removal (backward process) for image generation.

The forward phase embodies a Markov chain structure, $q(x_t|x_{t-1})$ progressively injecting Gaussian noise into the data ($x_0$) until it reaches a fully noise-perturbed image ($x_t$). Conversely, the backward process engages in a denoising mechanism, systematically removing the noise that exists in the previous timestep, adhering to the Markov Chain $p(x_{t-1}|x_t)$.

## Stable Diffusion

In the work by Rombach et al. (2022), the "Stable Diffusion" model tailors particularly for text-to-image synthesis tasks. The model operates by diffusing the latent vector representation of an image. It begins by receiving textual input, which is then transformed into a text embedding via the frozen CLIP text encoder (Radford et al. 2021). Subsequently, a text-conditional latent U-Net iteratively denoises the latent vector in a manner conditioned on the generated text embedding. Finally, a Variational Autoencoder (VAE) (Kingma and Welling 2013) decodes this latent vector, producing the corresponding image.

# Word-level Duplication

Carlini et al. (2023) introduces a definition of memorization wherein an example $x$ is considered extractable from a diffusion model $f_\theta$ if, without using $x$ as an input, there exists an efficient algorithm $A$ such that $\hat{x} = A(f_\theta)$ satisfies the condition $l(x, \hat{x}) \leq \delta$. This definition emphasizes replications that produce images nearly identical to the original. However, our focus is on a broader understanding of memorization, which we term *partial replication*. This pertains to specific objects or features within an image. Carlini's metric might not always capture such memorization; it may indicate a high $l_2$ distance even when recognizable memorization exists within images.

Somepalli et al. (2023b) investigates a broader scope of duplication in the LAION dataset, covering more cases than previous studies. They consider both caption and image duplications and even delve into partial caption duplication. However, there are concerns with their approach. They curated two subsets from the LAION dataset for fine-tuning the Stable Diffusion model. Fine-tuning the Stable Diffusion on a subset of its original pre-training dataset might result in increased unintended memorization.

In text-conditioned diffusion models, we believe that text plays a crucial role. Given this perspective, concerns should predominantly revolve around text-conditioned memorization in these applications. While image duplications might exist in the dataset, without a connection between the text and the image, it is improbable that related replication would emerge during inference time when supplying our prompt. This observation leads us to consider more realistic types of replication.

Unlike prior research, our focus is on *word-level duplication*. Specifically, we aim to discern any associations between keywords and images in duplications. We question whether certain sets of keywords and images are consistently replicated within the dataset. In this context, captions don't necessarily exhibit high semantic similarity; they might only share common keywords. Consequently, during inference, when the model encounters these specific keywords in combination, it might attempt to reproduce the corresponding features or objects observed during training. In our experimental results, we further explore this type of duplication through a detailed case study using the LAION dataset.

**A more realistic approach to defining memorization:** Previous studies have often relied on a single random initialization for generations (Somepalli et al. 2023b). However, irrespective of the memorization definition employed, we argue that a more realistic examination involves using multiple random initializations. Essentially, in practical settings, concerns about memorization and replication arise if the model consistently generates the same feature, object, or even the entire image across different initializations. Thus, assessing memorization or replication based on a single seed might not provide a comprehensive understanding.

# Object-level Duplication

In this section, we introduce a distinct type of duplication termed *object-level duplication*. This occurs when a pair consisting of specific objects in an image and certain keywords in the corresponding text is duplicated in the training dataset, even if the object's name does not appear in the text. Such duplication can lead to the replication of these specific objects during inference when the related keywords are present in the prompt. This pattern of replication raises various trustworthiness concerns, notably privacy and fairness. Essentially, it implies that the model persistently generates specific objects, irrespective of their mention or absence in the user-provided input, which may not align with user expectations or intentions.

A plausible explanation for this phenomenon might be the concealed correlation between certain keywords and objects within an image. That is, while entire images may not be duplicated in the training dataset, specific objects might frequently appear in images associated with captions containing a particular word. We delve into this phenomenon with a dedicated case study in the experiments section.

Table 1: Largest clusters with corresponding frequent words and their frequencies.

| Cluster ID | Keywords with Frequencies |
|---|---|
| 1 | **van**: 3061, **gogh**: 3042, **night**: 2841, **starry**: 2806 |
| 2 | **van**: 1950, **gogh**: 1937, **vincent**: 1374, **self-portrait**: 795, **portrait**: 674 |
| 3 | **van**: 1839, **gogh**: 1833, **almond**: 1764, **vincent**: 1201, **tree**: 1129, **blossoming**: 1003 |
| 4 | **van**: 1725, **gogh**: 1715, **sunflowers**: 1549, **vincent**: 1110, **vase**: 601 |
| 5 | **van**: 1628, **gogh**: 1622, **terrace**: 1477, **cafe**: 1313, **vincent**: 1135, **night**: 1034, **arles**: 530 |
| 6 | **van**: 1035, **gogh**: 1032, **night**: 955, **starry**: 925, **rhone**: 862, **vincent**: 597 |
| 7 | **van**: 906, **gogh**: 899, **irises**: 807, **vincent**: 586 |

## Experimental Results

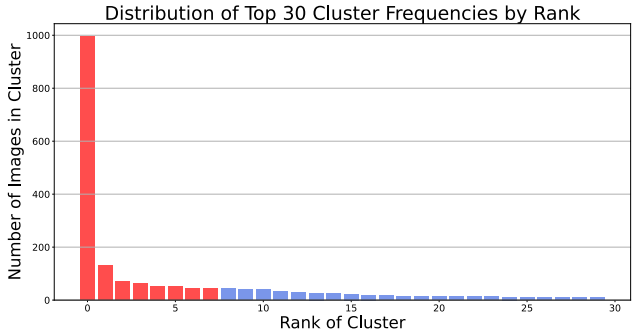



Figure 1: Frequency distribution of samples containing the words ”almond” and ”blossoming”. The red bars represent clusters related to the Van Gogh case.

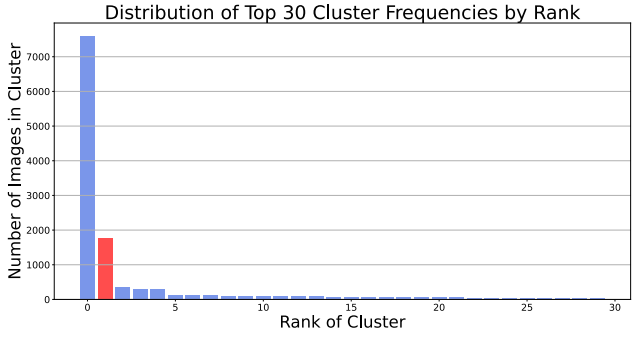



Figure 2: Frequency distribution of samples containing the word ”sunflowers”. The red bars represent clusters related to the Van Gogh case.

In this section, we present two case studies, each corresponding to one of the types of duplication previously discussed, and incorporate multiple examples within each study. For all experiments, we utilize the LAION-400M (Schuhmann et al. 2021), a subset of the larger LAION-5B (Schuhmann et al. 2022) dataset. This subset was chosen for its manageability in terms of scale. The experiments were conducted using the Stable Diffusion v1.4 model, which was trained on the LAION-5B dataset.

### Case Study 1: Van Gogh

In our initial case study, we delve into word-level memorization. For this purpose, we focus on samples with captions containing the term “Van Gogh”. Approximately 90,000 samples have this term in their captions. We proceeded to exclude samples with invalid URLs. Additionally, considering the text encoder of the CLIP model accepts text no longer than 77 tokens, samples with captions surpassing this token count were also omitted. Following these filtering steps, we were left with roughly 70,000 samples. Moreover, we obtained the image embeddings of these samples using the image encoder of the CLIP model.

In the next step, we cluster the image embeddings to identify sets of nearly identical images, utilizing the cosine similarity metric. Clusters are then ordered based on their size, and within each cluster, we pinpoint the most frequent words. It should be noted that the largest cluster, comprised of irrelevant images not closely related to others, has been omitted from our analysis. Table 1 presents the largest clusters along with their corresponding frequent words.

Now, we demonstrate how these keywords influence the generated images in each cluster. For each set of keywords, we consider the following captions:

- A caption composed solely of the keywords.
- A short relevant caption that includes the keywords.
- A long relevant caption that includes the keywords.
- An irrelevant caption that includes the keywords.
- A long caption excluding the term “van gogh”.

We obtain all of these captions using *ChatGPT* (OpenAI 2023). All captions and their corresponding generated images for cluster 1 are illustrated in Fig. 3. To better illustrate the concept of replication, for each prompt, we generate 500 images using different random initializations. Additionally, we present examples demonstrating varying levels of similarity to the original images in the training dataset. Furthermore, for each cluster, we establish a unique threshold for image similarities to determine the percentage of generations that are similar to the original images in the training dataset. This threshold varies among clusters and requires manual setting based on the specific characteristics of each cluster.

As demonstrated in Fig. 3, the experiment starts with a brief prompt and progresses to longer, more diverse captions. Regardless of the textual variations, the images con-

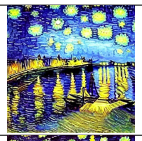
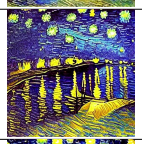
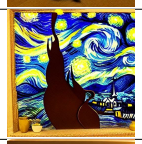
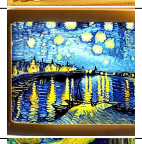
| Prompt | Text Similarity | Image Sim > 0.83 (%) | Example Generated Images | | | | Original Image |
|---|---|---|---|---|---|---|---|
| | | | > 0.85 | 0.80-0.85 | 0.70-0.80 | <0.70 | |
| Van Gogh starry night | 1.0000 | 64.6% |  |  |  |  |  |
| A starry night landscape filled with Van Gogh's vibrant colors. | 0.8279 | 56.8% |  |  |  |  |  |
| The starry night swirls in Van Gogh's signature style above a silent city, where the streets hum softly with the memory of daylight. | 0.7323 | 41.4% |  |  |  |  |  |
| A surreal scene of starry-eyed robots painting Van Gogh masterpieces during a power outage at night. | 0.5690 | 18.6% |  |  |  |  |  |
| Under this starry night, my rubber wader plotted world domination. | 0.4621 | 35.6% |  |  |  |  |  |

Figure 3: Captions containing the terms "Van Gogh", "starry" and "night," alongside their respective generated images for various seed values.

sistently maintain the style and elements of the original artworks. In the fourth example, even with "starry" and "night" separated, the images still jointly represent these themes. Intriguingly, the final caption omits "Van Gogh," yet his unique style is unmistakably captured in the images. Additionally, we calculate the cosine similarity between the given prompt and the closest text in the training dataset using CLIP's text encoder embeddings.

Besides the cluster whose examples are shown in Fig. 3, there is another cluster with intriguing outcomes. In Cluster 3, shown in Table 1, the key terms include "van gogh", "almond", and "blossoming". All captions and their corresponding generated images for this cluster are illustrated in Fig. 6 in the Appendix. The last example in Fig. 6 illustrates that even without explicitly mentioning "van gogh", the generated images bear a resemblance to those in the training dataset associated with Van Gogh's works. Moreover, you can find the captions and corresponding generated images for cluster 4 in Fig. 7 in the Appendix.

To understand this occurrence, we analyzed how frequently the words "almond" and "blossoming" are included in captions with "van gogh". By filtering out the dataset for captions with "almond" and "blossoming," we then clustered the images using image embeddings. It emerged that the dominant clusters, which are connected to Van Gogh's works, accounts for around 52% of the entries with these two descriptive words.

**Frequency matters.** Two main factors influence the likelihood of training image replication during inference. The first factor is the frequency of certain keywords within the dataset. Our observations indicate that images are more likely to replicate when associated with frequently occurring keywords. For instance, the words "starry night" and "almond blossoming" alongside "Van Gogh" have a higher propensity for replication.

However, frequency alone is not the only determinant. Another influential factor is the initial clustering of the dataset. When clustering is performed on images with specific keywords, such as "almond" and "blossoming," without including "Van Gogh," we find that the largest clusters still pertain to Van Gogh's works, representing about 52% of the samples. Nonetheless, a significant 48% of the clusters are unrelated. This distribution suggests that keyword frequency in the training set can predict model replication behavior to some extent. The keyword "sunflower" further exemplifies this; despite its frequent association with Van Gogh, it constitutes only 2% of the clusters when we consider only "sunflower" in the dataset. This underscores why Van Gogh's art style may not be replicated unless his name is explicitly mentioned. Fig. 1 illustrates the distribution size of the 30 largest clusters when we cluster images of samples whose captions contain the words "almond" and "blossoming". Fig. 2 shows the same thing for the word "sunflower".

## Case Study 2: Astronaut

In this section, we explore the concept of object-level replication through a focused case study. Object-level replication refers to the phenomenon where specific objects frequently appear in images despite their absence from the associated textual prompts. This implies a strong correlation between

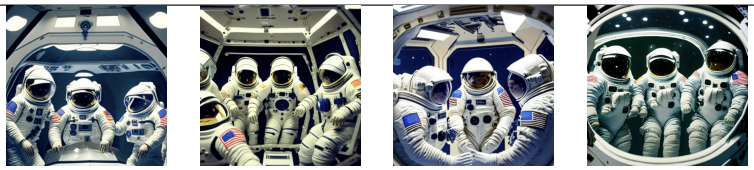
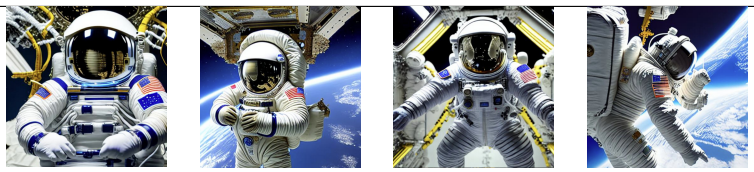
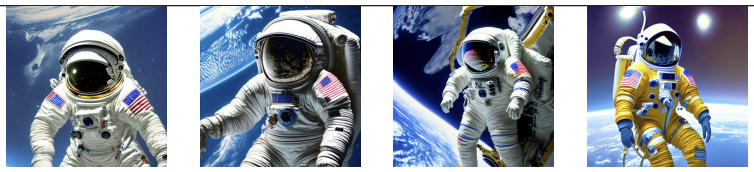
| Prompt | Example Generated Images | US Flag % |
|---|---|---|
| A child astronaut riding a dog through space |  | 24% |
| A group of astronauts in training inside a mock spacecraft. |  | 47% |
| A Chinese astronaut performing a spacewalk from the shuttle to test new satellite repair technologies. |  | 63.4% |
| An astronaut wearing a Russian Orlan spacesuit during a spacewalk. |  | 97.8% |

Figure 4: Prompts and their corresponding generated images with the percentage of images containing the US flag.

certain keywords and the recurrent visual elements within the dataset. To examine this phenomenon, we concentrate on samples from the LAION dataset containing the keyword "astronaut." We apply the same methodological framework as our initial case study to curate this subset of the dataset and to generate the corresponding image embeddings. This process resulted in approximately 48,000 samples, offering a substantial base for our investigation into keyword-object correlation. Fig. 5 presents some of these training samples whose captions contain the word "astronaut" and the corresponding images feature the US flag.

In this case study, our attention is on the US flag. An analysis of approximately 1000 training data samples with captions mentioning "astronaut" revealed that 10% included images of the US flag, even when the terms "US" or "flag" were not specified. To further explore this phenomenon, we first employed ChatGPT to craft a series of random prompts that include the word "astronaut." We then used these prompts to generate images with the Stable Diffusion model, which led to a frequent replication of the US flag in the output. Note that, due to the low quality of generation of the pre-trained Stable Diffusion model, we fine-tuned the model on a small dataset of prompts and corresponding high-resolution generated images from the Midjourney API to enhance the quality of the generated examples. Fig. 4 displays the ChatGPT-generated prompts and corresponding images from the Stable Diffusion model. By generating 500 images using varied random seeds, we assess the model's tendency to replicate the US flag from the prompt. Subsequently, we calculate and report the percentage of images featuring the US flag.

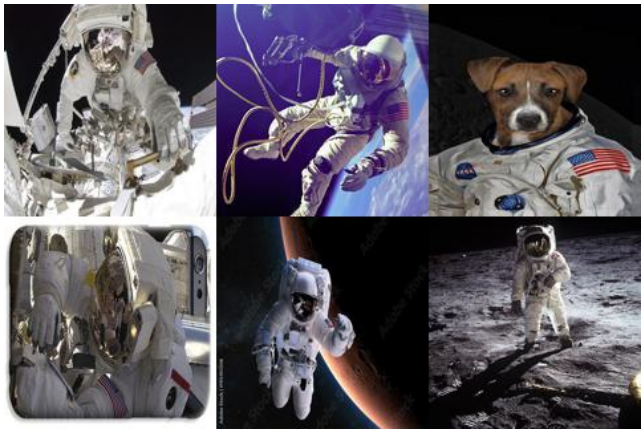


Figure 5: A collection of images from the LAION 400M training dataset showcasing astronauts with the US flag.

## Future Directions

Although our investigation is focused on two specific case studies, we have demonstrated the occurrence of word-level duplication in Stable Diffusion models. For future work, we propose conducting broader experiments within the word-level duplication context and undertaking a more comprehensive analysis. Additionally, developing new mitigation techniques that reduce memorization while preserving model utility is of paramount importance. The replicated features identified in our study also pose potential privacy risks, potentially making the models susceptible to various

attacks, including membership inference and backdoor attacks. Addressing these concerns will be a critical aspect of future research.

## Conclusion

Duplication in training data is a key contributor to memorization in generative models. This paper identifies two types of duplication leading to replication at inference. We investigated these through two LAION dataset case studies. Our work emphasizes the importance of vigilance against diverse duplication forms in training data and the need for effective mitigation strategies. It is our hope that this work will inspire more conscientious data curation and lead to the development of both powerful and privacy-preserving generative models.

## Acknowledgements

The work was supported in part by the NSF grant 2131910.

## References

Biderman, S.; Prashanth, U. S.; Sutawika, L.; Schoelkopf, H.; Anthony, Q.; Purohit, S.; and Raf, E. 2023. Emergent and predictable memorization in large language models. *arXiv preprint arXiv:2304.11158*.

Carlini, N.; Hayes, J.; Nasr, M.; Jagielski, M.; Sehwag, V.; Tramer, F.; Balle, B.; Ippolito, D.; and Wallace, E. 2023. Extracting training data from diffusion models. In *32nd USENIX Security Symposium (USENIX Security 23)*, 5253–5270.

Carlini, N.; Ippolito, D.; Jagielski, M.; Lee, K.; Tramer, F.; and Zhang, C. 2022. Quantifying memorization across neural language models. *arXiv preprint arXiv:2202.07646*.

Carlini, N.; Liu, C.; Erlingsson, Ú.; Kos, J.; and Song, D. 2019. The secret sharer: Evaluating and testing unintended memorization in neural networks. In *28th USENIX Security Symposium (USENIX Security 19)*, 267–284.

Carlini, N.; Tramer, F.; Wallace, E.; Jagielski, M.; Herbert-Voss, A.; Lee, K.; Roberts, A.; Brown, T.; Song, D.; Erlingsson, U.; et al. 2021. Extracting training data from large language models. In *30th USENIX Security Symposium (USENIX Security 21)*, 2633–2650.

Ho, J.; Jain, A.; and Abbeel, P. 2020. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33: 6840–6851.

Kim, G.; and Ye, J. C. 2021. Diffusionclip: Text-guided image manipulation using diffusion models.

Kingma, D. P.; and Welling, M. 2013. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.

Lee, J.; Le, T.; Chen, J.; and Lee, D. 2023. Do language models plagiarize? In *Proceedings of the ACM Web Conference 2023*, 3637–3647.

Midjourney. 2022. Midjourney.com.

Nichol, A.; Dhariwal, P.; Ramesh, A.; Shyam, P.; Mishkin, P.; McGrew, B.; Sutskever, I.; and Chen, M. 2021. Glide: Towards photorealistic image generation and editing with text-guided diffusion models. *arXiv preprint arXiv:2112.10741*.

OpenAI. 2023. ChatGPT on OpenAI Chat. Available at: https://chat.openai.com.

Radford, A.; Kim, J. W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, 8748–8763. PMLR.

Ramesh, A.; Dhariwal, P.; Nichol, A.; Chu, C.; and Chen, M. 2022. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125*, 1(2): 3.

Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; and Ommer, B. 2022. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 10684–10695.

Schuhmann, C.; Beaumont, R.; Vencu, R.; Gordon, C.; Wightman, R.; Cherti, M.; Coombes, T.; Katta, A.; Mullis, C.; Wortsman, M.; et al. 2022. Laion-5b: An open large-scale dataset for training next generation image-text models. *Advances in Neural Information Processing Systems*, 35: 25278–25294.

Schuhmann, C.; Vencu, R.; Beaumont, R.; Kaczmarczyk, R.; Mullis, C.; Katta, A.; Coombes, T.; Jitsev, J.; and Komatsuzaki, A. 2021. Laion-400m: Open dataset of clip-filtered 400 million image-text pairs. *arXiv preprint arXiv:2111.02114*.

Shi, Z.; Zhou, X.; Qiu, X.; and Zhu, X. 2020. Improving image captioning with better use of captions. *arXiv preprint arXiv:2006.11807*.

Shokri, R.; Stronati, M.; Song, C.; and Shmatikov, V. 2017. Membership inference attacks against machine learning models. In *2017 IEEE symposium on security and privacy (SP)*, 3–18. IEEE.

Sohl-Dickstein, J.; Weiss, E.; Maheswaranathan, N.; and Ganguli, S. 2015. Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning*, 2256–2265. PMLR.

Somepalli, G.; Singla, V.; Goldblum, M.; Geiping, J.; and Goldstein, T. 2023a. Diffusion art or digital forgery? investigating data replication in diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 6048–6058.

Somepalli, G.; Singla, V.; Goldblum, M.; Geiping, J.; and Goldstein, T. 2023b. Understanding and Mitigating Copying in Diffusion Models. *arXiv preprint arXiv:2305.20086*.

Webster, R.; Rabin, J.; Simon, L.; and Jurie, F. 2023. On the De-duplication of LAION-2B. *arXiv preprint arXiv:2303.12733*.

Zhang, C.; Bengio, S.; Hardt, M.; Recht, B.; and Vinyals, O. 2021. Understanding deep learning (still) requires rethinking generalization. *Communications of the ACM*, 64(3): 107–115.

## Appendix

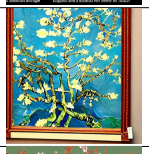
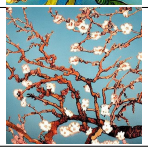
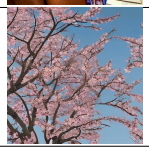
| Prompt | Text Similarity | Image Sim > 0.70 (%) | Example Generated Images | | | Original Image |
|---|---|---|---|---|---|---|
| | | | > 0.7 | 0.65-0.7 | <0.65 | |
| Van Gogh almond blossoming | 1.0000 | 87.2% |  |  |  |  |
| Van Gogh's vision of almond trees blossoming in the early spring light. | 0.8279 | 93% |  |  |  |  |
| Through Van Gogh's eyes, the blossoming almond branches are immortalized, their fleeting beauty captured in swirls of color and light. | 0.7323 | 90.4% |  |  |  |  |
| In a cozy corner of the library, a Van Gogh anthology sits open to a painting of almond trees blossoming, inspiring daydreams of spring. | 0.5690 | 37% |  |  |  |  |
| A digital art piece animating the life cycle of almond blossoming, each frame a celebration of growth and artistry. | 0.4621 | 5.4% |  |  |  |  |

Figure 6: Captions containing the terms "Van Gogh", "blossoming", and "almond", alongside their respective generated images for various seed values.

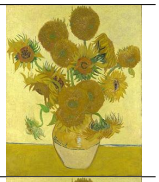
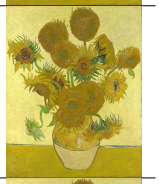
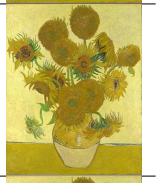
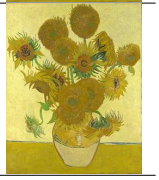
| Prompt | Text Similarity | Image Sim > 0.85 (%) | Example Generated Images | | | Original Image |
|---|---|---|---|---|---|---|
| | | | > 0.85 | 0.75-0.85 | <0.75 | |
| Van Gogh sunflowers | 1.0000 | 83.4% |  |  |  |  |
| A clay pot painted with sunflowers, like Van Gogh's art. | 0.8279 | 53.2% |  |  |  |  |
| Design a poetic intersection of nature and artistry, with sunflowers in full bloom amidst a scattering of Van Gogh's paint tubes and brushes on a wooden artist's table | 0.7323 | 42.4% |  |  |  |  |
| The new video game features a level where you collect sunflowers to unlock a character called Van Gogh. | 0.5690 | 25.8% |  |  |  |  |

Figure 7: Captions containing the terms "Van Gogh" and "sunglowers", alongside their respective generated images for various seed values.