

Not Every Image is Worth a Thousand Words: Quantifying Originality in Stable Diffusion

Anonymous Authors¹

Large-scale Text-to-Image (T2I) Generative Diffusion models have revolutionized our ability to generate and manufacture visual content using natural language descriptions. T2I models, as their name suggests, are designed to produce images given a textual prompt. Distinctively from a search-engine, T2I are not meant to find an existing image that fits a certain description, but they are supposed to *generate* novel content that fits the description of the text. Nevertheless, even though generation of new content is their defining trait, quantifying originality remains a formidable challenge both in practice as well as in theory.

This challenge is not solely scholastic, and arises in the context of legal concerns surrounding copyright laws, where T2I models, trained on expansive datasets like LAION-5B (Schuhmann et al., 2022) that include copyrighted materials, are often at the center of infringement accusations. Here too, quantifying originality poses a challenge as copyright law only protects the aspects of expressive works deemed *original* by the judiciary (Harper & Row, Publishers, Inc. v. Nation Enterprises, 1985; U.S.C, 1990), where originality necessitates a minimal degree of creativity and authorship (Feist Publications, 1991).

In turn, methodologically sound methods for demonstrating creativity and originality in a T2I model become a pressing matter. Traditional strategies often formalize the problem of not-copying as a form of memorization constraint that inhibits overfitting of the data (Carlini et al., 2023; Bousquet et al., 2020; Vyas et al., 2023). This is also highlighted in the recently implemented EU AI Act, which mandates the disclosure of training data (Institute for Information Law (IViR), 2023) that requires greater transparency in the operation and training of these models. However, regulating memorization is not necessarily aligned with the purpose of copyright law (Elkin-Koren et al., 2023), can be overly restrictive, and also poses computational as well as statistical challenges (Feldman, 2020; Feldman & Zhang, 2020; Attias

¹Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

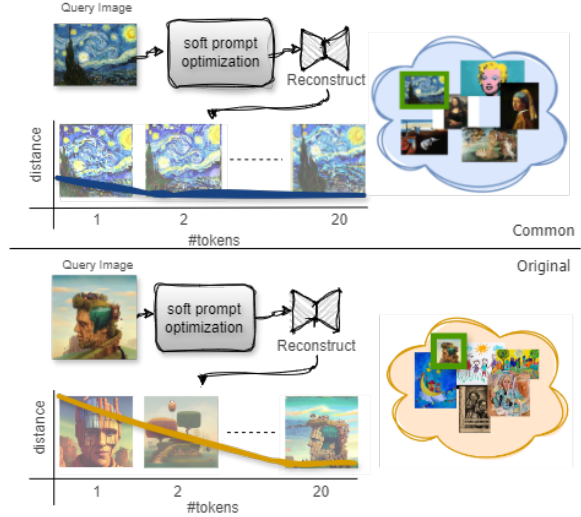


Figure 1. Illustration of our approach for measuring image originality using multi-token textual inversion. Original images require more tokens for accurate reconstruction, while common images like Van Gogh’s “Starry Night” need only one token.

et al., 2024; Livni, 2024; Zhang et al., 2016).

In this paper we consider an alternative viewpoint. We investigate whether T2I models can, in fact themselves, be utilized to discriminate between generic and original content. Towards this goal, we propose a quantitative framework that assesses the originality of images based on the model’s familiarity with the training data. We implement our framework concretely and provide real-world data experiments, that demonstrate the potential of T2I models in identifying originality in output content. We believe that our framework can be harnessed to build further metrics for originality and genericity, which in turn can be used to audit the utility of generative models, and hopefully be used to analyze originality in real-world images.

We introduce our conceptual framework to quantitatively measure originality or genericity of images. Inspired by the theoretical work of Scheffler et al. (2022), we look at the complexity of description as a measure of originality. The working hypothesis is that common works are easier to describe than original works. Unlike Scheffler et al. (2022) that builds on the notion of Kolmogorov complexity, here we utilize the fact that a T2I model was trained on a large















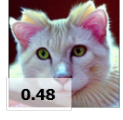


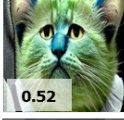


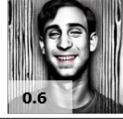



Domain	Query Image	1 token	2 tokens	3 tokens	4 tokens	5 tokens
Houses		 0.71	 0.48	 0.56	 0.38	 0.32
Art		 0.63	 0.53	 0.57	 0.52	 0.52
Animals		 0.6	 0.48	 0.41	 0.33	 0.52
People		 0.7	 0.6	 0.83	 0.7	 0.39

Figure 2. Qualitative results for reconstructing original images from various domains using multi-token textual inversion, demonstrating that for original images, more tokens improve capturing additional details of the query image. The average DreamSim score for each experiment is depicted at the bottom of each representative image.









	Houses	Art	Animals	People
Query Image				
Single Token Reconstruction	 0.31	 0.39	 0.28	 0.43

Figure 3. Qualitative results for reconstructing common images from various domains using multi-token textual inversion, demonstrating that for common images, a single token can reach high reconstruction capabilities. The average DreamSim score for each experiment is depicted at the bottom of each representative image.

corpus of real-world data and therefore we test whether common traits can be generated with shorter texts (Gal et al., 2022). In turn, by applying textual inversion techniques, we evaluate the extent to which a concept is familiar to the model, and thus, potentially unoriginal.

We validate our approach with empirical experiments utilizing the widely used Stable Diffusion model. Experiments employ both textual inversion and DreamSim (Fu et al., 2023) to analyze the correlation between the ease of concept recreation measured by the number of tokens needed and the originality of the images relative to the training dataset. Our experiments reaffirm and validate that embracing rather than avoiding memorization might enable generative models to produce more innovative and diverse content.

Overall, we introduce to the study of originality and copyright in generative models a new technique to identify genericity. In the full version of the paper, We elaborate on those

experiments by custom trained models with controlled synthetic data, and evaluate the overall genericity abilities of T2I stable diffusion models.

Method Our approach builds on the textual inversion technique by (Gal et al., 2022), extending it to use multiple tokens to enhance the interpretability of text-to-image (T2I) models and focus on image originality. Our experiments show that the number of tokens required for reconstruction correlates with image originality, serving as a measure of originality. By representing concepts with multiple tokens, we achieve a more detailed latent space representation, facilitating deeper exploration of T2I models. We employ DreamSim to assess visual quality, comparing generated images to the query image. Our method, tested in both synthetic and real-world settings, ensures in-domain generation and originality by using different criteria for each context. This approach, depicted in Fig. 1, enhances the assessment of image originality and the interpretability of T2I models, providing a comprehensive framework for evaluating and understanding originality for generated images.

Results In this section, we present our experimental results, showing that original images require more tokens for accurate representation. Qualitative examples for reconstruction using textual inversion are shown in Figure 2 for original images (labeled by a human expert) and in Figure 3 for common images. As observed, semantic preservation improves with more tokens for original content and is already high with the first token for common content. This is further supported by the DreamSim scores, which are significantly lower for the original images.

Impact Statement

This paper presents work whose goal is to advance the field of Machine Learning by introducing a framework for quantifying originality in text-to-image generative diffusion models. The potential broader impact of this work includes the following:

Ethical Aspects Our research addresses the challenge of quantifying originality, which has significant implications for copyright laws and the protection of creative works. By providing a methodology to assess the originality of generated images, we aim to contribute to a fairer and more transparent use of generative models in creative industries. This could help mitigate legal disputes related to copyright infringement and ensure that the rights of original content creators are respected.

Future Societal Consequences The ability to quantify originality in generated images could enhance the deployment of generative models in various fields, including art, design, and entertainment, by fostering trust and accountability. It can also encourage the development of new creative tools that assist artists in generating unique content while respecting intellectual property rights.

While the primary goal of this work is to advance the field of Machine Learning, we believe that our contributions to the understanding of originality and creativity in generative models will have a positive societal impact by promoting ethical use and fostering innovation.

References

- Attias, I., Dziugaite, G. K., Haghifam, M., Livni, R., and Roy, D. M. Information complexity of stochastic convex optimization: Applications to generalization and memorization. *arXiv preprint arXiv:2402.09327*, 2024.
- Bousquet, O., Livni, R., and Moran, S. Synthetic data generators—sequential and private. *Advances in Neural Information Processing Systems*, 33:7114–7124, 2020.
- Carlini, N., Hayes, J., Nasr, M., Jagielski, M., Sehwal, V., Tramer, F., Balle, B., Ippolito, D., and Wallace, E. Extracting training data from diffusion models. In *32nd USENIX Security Symposium (USENIX Security 23)*, pp. 5253–5270, 2023.
- Elkin-Koren, N., Hacohen, U., Livni, R., and Moran, S. Can copyright be reduced to privacy? *arXiv preprint arXiv:2305.14822*, 2023.
- Feist Publications. 499 u.s. 340. pp. 345, 1991.
- Feldman, V. Does learning require memorization? a short tale about a long tail. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, pp. 954–959, 2020.
- Feldman, V. and Zhang, C. What neural networks memorize and why: Discovering the long tail via influence estimation. *Advances in Neural Information Processing Systems*, 33:2881–2891, 2020.
- Fu, S., Tamir, N., Sundaram, S., Chai, L., Zhang, R., Dekel, T., and Isola, P. Dreamsim: Learning new dimensions of human visual similarity using synthetic data. *arXiv preprint arXiv:2306.09344*, 2023.
- Gal, O., Patashnik, O., Maron, H., Chechik, G., and Cohen-Or, D. Image specific fine-tuning of text-to-image diffusion models. *arXiv preprint arXiv:2208.01618*, 2022.
- Harper & Row, Publishers, Inc. v. Nation Enterprises. 471 u.s. 539. pp. 547, 1985.
- Institute for Information Law (IViR). Generative ai, copyright and the ai act. Kluwer Copyright Blog, May 2023. URL <https://copyrightblog.kluweriplaw.com/2023/05/09/generative-ai-copyright-and-the-ai-act/>. Retrieved March 6, 2024.
- Livni, R. Information theoretic lower bounds for information theoretic upper bounds. *Advances in Neural Information Processing Systems*, 36, 2024.
- Scheffler, S., Tromer, E., and Varia, M. Formalizing human ingenuity: A quantitative framework for copyright law’s substantial similarity. In *Proceedings of the 2022 Symposium on Computer Science and Law*, pp. 37–49, 2022.
- Schuhmann, C., Beaumont, R., Vencu, R., Gordon, C., Wightman, R., Cherti, M., Coombes, T., Katta, A., Mullis, C., Wortsman, M., et al. Laion-5b: An open large-scale dataset for training next generation image-text models. *Advances in Neural Information Processing Systems*, 35: 25278–25294, 2022.
- U.S.C. 17 U.S.C. § 102(a). 1990.
- Vyas, N., Kakade, S., and Barak, B. Provable copyright protection for generative models. *arXiv preprint arXiv:2302.10870*, 2023.
- Zhang, C., Bengio, S., Hardt, M., Recht, B., and Vinyals, O. Understanding deep learning requires rethinking generalization. *arXiv preprint arXiv:1611.03530*, 2016.