Protecting Art against Generative AI: Threat Perceptions and Countermeasure Advice on TikTok and YouTube

Dana Rohmann
Paderborn University

Rachel Gonzalez Rodriguez
Paderborn University

Anna Lena Rotthaler
Paderborn University

Harshini Sri Ramulu
Paderborn University

Yasemin Acar
Paderborn University

Abstract

Artificial Intelligence (AI) image generators are commonly trained on publicly available digital artworks without creators' consent, enabling style mimicry of individual artists and raising concerns about intellectual property protection. To investigate advice related to AI image generators, we explore and thematically analyze 94 videos by 82 different accounts sourced from TikTok and YouTube addressing protection methods for sharing art online. We find that artists frame data scraping as the fundamental mechanism behind three categories of perceived threats: style mimicry directly harms individual artists; on a structural level, professional artists are harmed by a loss of income; and on a societal level, AI-generated images can be used for malicious purposes and the fabrication of creative effort. As countermeasures, we find that "AI disturbance" is prevalent advice with viral appeal in short-form content, but lacks explanation and reliable sources. In contrast, academic Glaze and Nightshade are better explained and sourced, but face some accessibility barriers [8, 9]. Creators also discuss non-technical countermeasures: switching social media platforms, altering posting practices, and adjusting data and privacy settings. We recommend that security experts engage with popular content formats to improve advice quality, and call for consent-based frameworks for AI training data developed in collaboration with artists.

1 Introduction

Artificial Intelligence (AI) image generators have become widely known since 2022 and generate images by processing text-based prompts [1, 5]. More companies try to supplement or replace human work with AI-generated content, raising concerns about copyright [6], reduction of artist wages [2] and non-disclosure of training data [3]. Throughout this paper, we use the term *"AI art theft"* to refer to the non-consensual use of artwork in training AI. The commercialization of AI art and concurrent style mimicry might pose copyright violations. Based on moral foundations of copyright laws, this practice can be framed as theft [3, 4, 10]. In order to prevent publicly shared artworks from being used for AI, artists take self-implemented security measures to protect their artworks, through modified sharing behaviors and application of technical methods. Artists use social media platforms like TikTok and YouTube to convey strategies to protect their art against AI. In this study, through qualitative analysis of 94 advisory videos by 82 unique accounts published on TikTok and YouTube, we answer the following research questions:

RQ1: Analyzing Threat Models: *How do content creators portray threats posed by AI image generation for artists and their intellectual property?*

RQ2: Characterizing AI Protection Advice: *What countermeasures do content creators recommend, and what sources, if any, do they use to support their advice?*

2 Methods

Data collection. We systematically collected and analyzed 94 videos on TikTok and YouTube to understand the discourse on threat perceptions and countermeasures surrounding AI art theft. We utilized the search function on YouTube and TikTok to find advice for digital artists to safeguard their artwork using search terms and hashtags. We conducted multiple rounds of video collection, iteratively adding hashtags to our search terms each round and stopped when no more new videos surfaced. To ensure video recommendations unaffected from prior search history, we created completely new accounts for each platform at the start of data collection. We used a VPN configured to a US server location to collect videos in English and to ensure a consistent search context. A video needed

USENIX Symposium on Usable Privacy and Security (SOUPS) 2026.
August 23–26, 2026, Hannover, Germany.

to satisfy three inclusion criteria for analysis: (1) it is about preventing artworks from being used by AI, (2) it gives advice, and (3) it was uploaded after 31.07.2023—covering the release Glaze (August 2023) and the then latest Midjourney version (V6) . We collected data using 33 search queries in three iterations. During data collection, we evaluated over 1320 videos, and 98 videos met our inclusion criteria. We further excluded 3 out of 98 videos that were posted by content farms and one video because the creator appeared to be under the age of 18. The final dataset consists of 94 videos: 50 from TikTok and 44 from YouTube, of which 16 are YouTube Shorts.

Data analysis. Two researchers coded all 94 videos qualitatively using an inductive open coding approach. Both researchers double coded all videos iteratively until they resolved all disagreements in the codebook (Appendix A).

Ethical considerations. We acknowledge the public nature of social media, and follow ethical best practices for data collection and analysis, closely adhering to the Menlo Report principles [7]. To mitigate the risk of re-identification of content creators, we removed any Personally Identifying Information (PII) and paraphrased all quoted statements.

3 Key Findings

Threat framing. Non-consensual involvement in AI training datasets is the most discussed concern (54 videos). About half of the videos describe the subject as "theft" or terminology that implies theft (37 videos). We identify that companies and individuals who either scrape data for AI training or utilize generated images are considered threat actors.

Direct threats to individual artists. Style mimicry is an underlying threat that enables multiple harms. When an artist's name is used to generate artworks with style-specific properties, fans and potential customers can mistake the AI image for an authentic image by the mimicked artist. AI images are not broadly protected by copyright laws as of the time of writing, making AI-generated images a way to utilize an artist's work as copyright-free material.

Structural threats. Some creators specified the negative impact AI images have on the current artistic job market (14 videos). An issue lies within the combination of unauthorized data usage and the lack of credit and compensation. Some videos discuss copyright issues and argue that AI image generators infringe upon copyright protections (18 videos). Discourse about copyright infringement appears to be based on moral values and collective sentiments rather than legal evaluation.

Wider societal impacts. Video creators voiced their concern about using AI generation for malicious purposes like producing political propaganda in bulk and revenge images of women (11 videos). The underlying issue, as presented in four videos, is the sense of accomplishment from generating images despite the lack of effort and meaning behind the output.

Technical defenses. Multiple creators seem to reenact a trend called "AI disturbance" (35 videos), which uses visual noise overlays on artworks. "AI disturbance" advice does not provide explanations about how and why this protection method works. Most videos also do not provide any advice sources (23 videos). Glaze is the third most recommended protection method (32 videos); half of the videos that advise using Glaze mention the official website as source. Nightshade is the most discussed protection method (37 videos). While the videos correctly reference official resources for Nightshade and Glaze, a few creators misrepresent or exaggerate protection capabilities.

Behavioral defenses Protecting artworks by changing social media activities is recommended by several videos (17 videos). Especially Instagram (and Meta) are frequently criticized for allowing AI training with posted images. The platform Cara is promoted and praised for being seriously interested in protecting artists from exploitation by AI.

4 Discussion

Overall, Glaze and Nightshade are well established within the protection advice discourse, especially prominent in long-form videos on YouTube. Both are presented as secure protections. This ignores that both are controversial in the scientific discourse, and have successfully been attacked. This may mislead users into a false sense of safety when sharing artworks online.

"AI Disturbance" comes with more severe problems compared to Glaze and Nightshade. Most videos lack audio instructions, explanations, and a source. We assess this advice as unreliable and a potential risk for artists, because of lacking sources that instruct on the verification of correct execution and effectiveness of the protection.

We recommend that AI security experts should improve advice communication by creating advice content about validated protection methods, utilizing popular formats and platforms to reach a wide audience of non-expert artists. Companies, especially social media platforms, should only allow data scraping for AI training data with informed consent. We emphasize the importance of clear taxonomies for data that may be used for AI training as opposed to personal data.

References

[1] Eleonora Vilgia Putri Beyan, Anastasya Gisela Cinintya Rossy, et al. A review of ai image generator: influences, challenges, and future prospects for architectural field. *Journal of Artificial Intelligence in Architecture*, 2(1):53–65, 2023.

[2] Dani Di Placido. Coca cola's ai-generated ad controversy, explained. *Forbes*.

[3] JOSEPH SAVERI LAW FIRM. Midjourney lawsuit. https://regmedia.co.uk/2024/01/03/midjourney_lawsuit.pdf, 2023. Case 3:23-cv-00201-WHO Document 129 Filed 11/29/23, pp. 19-20.

[4] Trystan S Goetze. Ai art is theft: Labour, extraction, and exploitation: Or, on the dangers of stochastic pollocks. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, pages 186–196, 2024.

[5] James Hutson and Morgan Harper-Nichols. Generative ai and algorithmic art: disrupting the framing of meaning and rethinking the subject-object dilemma. *Global Journal of Computer Science and Technology: D*, 23(1), 2023.

[6] Reishiro Kawakami and Sukrit Venkatagiri. The impact of generative ai on artists. In *Proceedings of the 16th Conference on Creativity & Cognition*, pages 79–82, 2024.

[7] Erin E. Kenneally, David Dittrich, and Michael Bailey et al. The menlo report: Ethical principles guiding information and communication technology research, August 2012.

[8] Shawn Shan, Jenna Cryan, Emily Wenger, Haitao Zheng, Rana Hanocka, and Ben Y Zhao. Glaze: Protecting artists from style mimicry by {Text-to-Image} models. In *32nd USENIX Security Symposium (USENIX Security 23)*, pages 2187–2204, 2023.

[9] Shawn Shan, Wenxin Ding, Josephine Passananti, Stanley Wu, Haitao Zheng, and Ben Y. Zhao. Nightshade: Prompt-specific poisoning attacks on text-to-image generative models. In *2024 IEEE Symposium on Security and Privacy (SP)*, pages 807–825. IEEE.

[10] Eric Shoemaker. Is ai art theft? the moral foundations of copyright law in the context of ai image generation. *Philosophy & Technology*, 37(3):114, 2024.

[11] Warda Usman and Daniel Zappala. Sok: A framework and guide for human-centered threat modeling in security and privacy research. In *2025 IEEE Symposium on Security and Privacy (SP)*, pages 2697–2715. IEEE, 2025.

A Codebook

Code	Sub-codes
Threat Models	
General Threats of AI Generated Images	Negative environmental impact, Misinformation, Propaganda images, Harassment, AI art accusations, Biases and stereotypes, Fabricated effort
Devaluing Artists' Labor	Negative impact on job market, Decreasing job opportunities for artists, Decreasing entry level jobs, AI is cheaper than paying an artist, Companies using AI images, Less visibility for human art, Discouraging studying art
Damaging Artists' Reputation	False representation, Loss of authority over own art, Less commissions, Misuse of visual identity, Selling AI art forgery
Copyright Infringement	Artists unaware of rights, Need evidence of ownership, Feeling unprotected by laws, Not up to date with AI
Non-consensual Involvement	Scraping, Theft, Without compensation, Without credit, Without permission, Data privacy violations, Lack of transparent datasets
Style Mimicry	AI and human learning process comparison, Damaging artists' reputation, False representation, Hobby of copying art styles, Less commissions, Loss of authority, Visual identity, Selling AI art forgery
Glaze	Accessibility issues, Opacity recommendations, WebGlaze, Cloaking, Use cases, Technical explanation, Opinions: Visible artifacts; Barely noticeable changes; Looks interesting; Better suited for certain styles
Nightshade	Accessibility issues, Do not mention Nightshade in posts, Poisons one concept, Ethical considerations, Explaining sample size, Poison attacks, Rename files before posting, Explaining metadata tags, Disturb AI training models, Explaining use cases, Technical explanation, Opacity recommendations, Opinions: Visible artifacts; Barely noticeable changes; Looks interesting; Fear of false AI reports; Altering artist style; Expressing lack of knowledge
Miscellaneous Methods	Watermarks, Signatures, Change art styles
Countermeasures	
AI Disturbance Texture	Commonly reused pattern, Premade AI Disturbance pattern, Personalization, Generic image layering, Ibis-Paint Premium, Use letters as overlay, Opacity recommendations, Opinions: Distrust method; Express uncertainty about protection; Positive statements; Barely noticeable
Restricting Shared Content	Do not upload originals, Remove posted artworks, Upload angled images, Upload process videos, Delete your account, Delete posts, Reduce posting activity, Reduce resolution of images, Showcase artworks only by videos, Do not opt in to AI features
Avoid AI Friendly Platforms	Opting out, Opting out barriers, Automatic opt-ins, Named platform/company
Use AI Hostile Platforms	Named platform/company, Accessibility settings, Scraping can still happen, Use **Cara**: Glaze integrated; Limited audience; Without AI features; AI content is forbidden; No AI label for crawler
Taking Legal Actions	Sue because of Copyright Infringement, Consult a Lawyer, License your work, Take down request, Fight policy changes
Advocate for Human Art	Activism, Build a community, Share campaign poster, Increase media coverage, Notify original artist, Be critical about Reddit advice, Do not use AI generators
Advice Presentation	
Advice Source	Official publisher website, Paper, Social media, Other artists, Own experience as artist, No source provided, Lawyer, Online articles, Terms of service
Video Attributes	Step-by-Step Instructions, Self-experiment, Testing protection
Simplification	Without audio instructions, Without explanations, Inaccuracy

Table 1: Codebook.

B Threat Model Overview

Table 2: Threat model overview based on the human-centered threat modeling framework by Usman and Zappala [11]. The components *context*, *threats*, and *protective strategies* are derived directly from the 94 TikTok and YouTube videos in our dataset; *considerations* and *reflections* are based on our interpretation of the data.

Component	Factor	Description of Factor Manifestation in Data
Context	Risk Factors	Posting artworks publicly online; professional artists need online exposure as essential part for business; prominent artists at higher risk due to recognizable style and name association
	Unique Challenges	Less online exposure risks decrease of business opportunities; majority of artists are not technical/legal experts; moral values and current legal frameworks diverge
	Culture	Sharing artworks online is a norm in the artistic community; collective framing of AI training as "theft"; perceive style mimicry as copyright infringement; visual identity is part of personal identity
	Personal Identity & Perception	Prominent artists face more potential harms than average artists; artists perceive style choices as visual identity and part of their personal identity
	Technology Background	IT creators as minority in dataset; artists lack IT expertise
Threats	Concerns	Non-consensual data scraping; style mimicry; lack of legal safeguards
	Harms	Reputational damage; misuse of visual identity; less commissions/jobs; devaluing creative work
	Actors	Companies that scrape data for AI training data sets; users of AI image generators mimicking art styles; social media platforms with opt-out policies for data mining
	Trade-offs	Protection methods alter artwork aesthetics; Nightshade needs enough samples to be effective; Glaze does not protect in retrospect; online visibility in contradiction to data privacy
	Threat Appraisal	Average artists question personal risk, but are aware and still share protection methods; prominent artists feel threatened on a personal and professional level
Protective Strategies	Recovery	Signatures/watermarks to prove ownership for future legal action; take-down requests for AI images; retroactive dataset poisoning with Nightshade
	Information Seeking	Watching TikTok/YouTube tutorials; official Glaze/Nightshade resources; community advice; online articles
	Technical Defenses	Glaze; Nightshade; "AI Disturbance" overlays; signatures and watermarks
	Behavioral Defenses	Posting at angles; posting process videos instead of finals; avoiding social platforms that mine user data; switching to Cara
Considerations	Mapping	Glaze $\rightarrow$ style mimicry prevention; Nightshade $\rightarrow$ data poisoning; "AI Disturbance" $\rightarrow$ style mimicry prevention / decrease of AI output quality (vague, unverified); Watermarks/Signatures $\rightarrow$ proof of authorship; switching to Cara and posting alternative versions $\rightarrow$ scraping prevention
	Perceived Cost	Less freedom of artistic choices and self-representation online; hardware requirements for Glaze/Nightshade
	Coping Appraisal	Trend methods ("AI disturbance") without proof of efficacy; Nightshade needs large sample sizes; dissatisfaction with current legal safeguards; minor overestimation of protective quality of Nightshade and Glaze
Reflection	Efficacy	"AI Disturbance" perceived as useful but unverified; Glaze/Nightshade better evidenced but misconceptions exist; no information about changed posting habits; take-down requests unresolved
	Barriers	Accessibility issues of WebGlaze (invite-only); hardware requirements for Glaze/Nightshade; aesthetic tradeoffs; inaccessible opt-out settings; inconsistent AI disturbance instructions
	Advice	Artistic and IT-community advice on TikTok/YouTube; absence of official Glaze/Nightshade channels
	Visioning	Consent-based frameworks; legal reforms; compensation and credit for artists; opt-in systems; expert advice in popular content formats; prompt filter against style mimicry