Baltic J. Modern Computing, Vol. 13 (2025), No. 4, 778-805
https://doi.org/10.22364/bjmc.2025.13.4.03

Effectiveness of Image Protection Software Against Image Generation Tool Training

Valerija JANUSEVA, Solvita ZARINA

Faculty of Science and Technology, University of Latvia, Riga, Latvia

valerijajanuseva@gmail.com, solvita.zarina@lu.lv

ORCID 0009-0008-4753-767X, ORCID 0000-0001-8884-2971

Abstract. Unauthorised use of artworks in training image generation models poses a growing challenge for online copyright protection. Contemporary artists' work is used without their consent both by artificial intelligence (AI) companies and Internet users, creating an urgent need for effective protective measures. The article examines and compares software solutions designed to safeguard artworks from such unauthorised use, examining both technical effectiveness and the output's visual quality. We evaluate different image protection tools and their change intensity levels by applying them to illustrations created by one of the authors and then training generative models on both protected and unprotected versions. The resulting images are evaluated by a voluntary survey of 71 respondents including artists and artwork viewers (non-artists). The discussion and conclusions assess image protection software based on research findings, provide recommendations for artists, outline future research opportunities, and demonstrate that participation increased respondents' awareness of the importance of protecting artworks.

Keywords: Image Protection Software, Protection of Artworks, Adversarial Perturbations, Testing, AI Model Training, Image Generation

1. Introduction

Image generation using artificial intelligence (AI) models, which has become widely accessible with the release of tools like Midjourney and Stable Diffusion in 2022, is now commonly used for entertainment, personal, and professional purposes. However, significant ethical and legal issues have been discovered – the image generation models were trained on millions of visual data samples obtained from the Internet without the consent of their creators, thus negatively impacting artists' careers and infringing on their copyright (Heikkilä, 2022).

In response to unauthorised use of artworks, several image protection tools have been developed that introduce subtle visual changes into images, affecting the ability of AI models to mimic artistic styles. Three promising tools are Glaze, Nightshade, and Mist. Although previous research shows their potential and technical feasibility, it focuses on each tool's individual performance and uses a separate experiment design. Furthermore, not all experiments include digital illustrations and art, which artists often publish online, and which are particularly vulnerable to unauthorised use for AI training.

This motivates us to address these gaps by conducting comparative technical experiments, analysing the visual evaluation of the results, and investigating the impact of the software on a less studied medium of artwork creation. Thus, the aim of this article is to investigate and compare the effectiveness of such image protection software tools in limiting the ability of AI to replicate an artist's individual style. We hypothesise that training generative models on protected images negatively impacts their ability to mimic artists' style effectively and that the introduced changes do not affect the perceived visual quality of the artwork. Therefore, artists can be advised to use these tools to protect their works. Thus, the objectives of the research are to conduct an experiment by training small image generation models with both unprotected and protected image samples, and to compare and evaluate the effectiveness of protection software with the participation of artists and viewers of artworks in the study.

To test the hypothesis, we proceed as follows: (a) we review the current situation and provide an insight into the relevant literature on image generation and its principles, as well as on image protection software, (b) we conduct an experiment in which we train eight small-scale image generation models, one with unprotected and seven with protected datasets, each using different protection software and change intensity settings, (c) we evaluate the effectiveness of protection software through a user survey among 71 respondents including artists and viewers of artworks (non-artists) and compare the survey results with computational measurements of style similarity. Finally, based on the results of the experiment and the evaluations obtained, we compare the three tools and provide artists with recommendations on how to protect images, as well as discuss future research directions and acknowledge the survey's impact on raising artists' awareness on image protection.

2. Insight into image generation

2.1. Review of the current situation

In the past three years, the rapid growth of generative AI tools has taken the technology world by storm. It is now possible to type in a prompt with just a few words or phrases and, within seconds, generate a high-quality image that may be almost indistinguishable from authentic human-made photographs or works of art. "The obvious source of these systems' popularity is that they offer something entirely new: being able to generate an image just by describing it, without having to go to the trouble of learning a skill – such as illustration, painting or photography – to actually make it." (McCormack et al., 2023).

AI-generated content is now regularly featured on social media platforms, and it has even won prizes in art and photography competitions (Roose, 2022; Parshall, 2023). Large companies such as Coca-Cola have used generative AI for creating video adverts (Coca-Cola, 2024). The popular online platform for films and TV series, Netflix, plans to introduce AI-generated advertising in the middle of streaming from 2026 (Harding, 2025).

In March 2025, the company OpenAI released the ChatGPT model GPT-4o with capabilities to generate improved quality images in a "wide range of styles" (OpenAI, 2025). Internet users quickly realised that the model could convincingly replicate the style of animated films from the renowned Japanese animation studio, Studio Ghibli. This led

to a flood of images generated in the artistic style of the films on Instagram, Facebook, and X (formerly Twitter). The release of this model pushed the average number of weekly active ChatGPT users to over 150 million for the first time (Sriram, 2025). In Latvia, a Japanese cuisine restaurant Shōyu Ramen generated a Ghibli-like animated promotional reel that was published on its Instagram account (Shōyu, 2025).

Many ethical concerns surround these practices. The AI company Midjourney has published a list of 16,000 artists whose artworks were collected to train their image generation tool. The list includes not only historical artists but also contemporary illustrators, some of whom have worked for corporations like Nintendo and Hasbro (Ho, 2023). The company OpenAI itself admits that it is unable to develop its products without copyright infringement – otherwise, there would not be enough data to train its models sufficiently (Milmo, 2024).

The aforementioned Studio Ghibli has not yet publicly commented on the imitation of the artistic style of its films using models such as GPT-4o. The studio is known for its traditional and hand-drawn animation techniques. In 2016, Hayao Miyazaki, the studio's founder, criticised the proposal to develop and use a machine that draws like a human in filmmaking, saying: "We humans are losing faith in ourselves." (MPNFW, 2016). Given the studio's principles, it is difficult to imagine that they would willingly permit their works to be used in AI model training, but it cannot be ruled out in the absence of an official statement. In the meantime, Japanese politicians have begun to discuss the legal ramifications of this situation. While it would be the studio's own responsibility to initiate legal proceedings, the politicians have commented: "If AI-generated content is determined to be similar to or reliant on preexisting copyrighted works, then there is a possibility that it could constitute copyright infringement" (Mullicane, 2025).

Professional artists invest significant time in mastering various forms of art and developing distinctive artistic styles. Many are unhappy that their work is being used for AI training without their consent or financial compensation. Legal actions have been initiated against several companies, such as Stability AI, and Midjourney. Although the court initially dismissed the artists' claims, it recognised the possibility of illegal use of copyrighted material in 2024 after a re-filing with amended arguments (Porterfield, 2023; Cho, 2024).

The legal status of generative AI currently remains unresolved. Since image generation models do not store the actual dataset but convert it into model weights during training, it may be challenging to prove copyright infringements. Furthermore, not all major AI companies disclose the datasets they use for model training, and some, such as Stability AI, argue that their actions constitute pastiche rather than copyright infringement (Wyn Davies, 2024). Clear legal guidelines, therefore, still need to be established.

In 2024, the European Union (EU) introduced the AI Regulation 2024/1689, which will come into force in 2026. Article 105 of the Regulation acknowledges that the development of generative AI poses problems for artists, authors, and other creatives. Although generative AI tools are not classified as high-risk, they do belong to a category that is subject to transparency requirements. In 2026, generative AI companies will have to disclose that content is generated by AI and publish sufficiently detailed descriptions of training datasets containing copyrighted material (European Parliament, 2024). Although this does not completely solve the fundamental problem, it is a step towards protecting creators.

In the EU, works of art are automatically protected by copyright upon creation, and the copyright lasts for up to 70 years after the death of the author (WIPO, 2003). It is not mandatory to go through a formal application process, but optional registration is possible if desired (Your Europe, 2025). It should be noted that artificial images generated by AI tools are not currently protected by copyright. In the United States, courts have ruled in several cases that the author of an artwork must be a human in order to be protected by copyright (Brittain, 2023).

2.2. Principles of image generation

This section outlines how image generation models work, in order to provide an understanding of how exactly they interact with and learn from visual data.

AI models are neural networks consisting of multiple layers of artificial neural nodes that mimic the behaviour of biological neurons. These neural networks require large training datasets to learn to perform a variety of tasks as accurately as possible, including image generation. There are different methods for training neural networks.

Generative adversarial networks (GANs), developed in 2014, were among the first deep learning architectures capable of generating new images. GANs consist of two dynamically updated neural networks – a generator and a discriminator. The generator is trained on a dataset of images to create new artificial images with the aim of increasing the probability that the discriminator will make a mistake. The discriminator is a classifier that predicts whether the provided image is a real or an artificial one. The two networks compete during training until the generator is able to produce such convincing images that the discriminator struggles with classification. At this point, the GAN can be used for image generation tasks (Goodfellow et al., 2014).

A variational autoencoder (VAE) is another type of machine learning architecture that consists of two components – an encoder and a decoder. The encoder takes input data from a dataset and tries to understand its features. The data is compressed into the latent space, which is a low-dimensional space where only meaningful information about the input data is retained. Instead of outputting a single point of data, VAE outputs the standard distribution in the latent space, which shows how much the values can vary. The decoder selects a single value from the distribution and attempts to reconstruct the original input by generating new data. Incorporating variation into the process provides more diverse generation possibilities and ensures that the model does not simply memorise the original dataset (Bergmann and Stryker, 2024).

Diffusion models are currently the most commonly used architecture for image generation. To learn to generate new images, diffusion models gradually add Gaussian noise to the input image until it contains only noise that bears no resemblance to the original. The models then learn the structure of the input images by reversing the diffusion process - removing the noise in a structured way to gradually reveal more image features, e.g. eyes, lips, etc., in a photo of a human. After training, the model can generate images from randomly selected noise that are not direct copies of the images in the original dataset but are very similar in structure (Sohl-Dickstein, 2015). Diffusion models can be combined with large language models (LLMs) to create a guided diffusion model, e.g. text-to-image models such as Stable Diffusion and Midjourney.

Diffusion models offer higher quality and more stable results than GAN models, but they are computationally intensive and therefore much slower. Stable Diffusion models use a modified and improved approach - using the principles of VAEs, the diffusion process is implemented in latent space rather than in the pixel space of the image. Such latent diffusion models require fewer computational resources while providing higher output quality (Rombach et al., 2022).

2.3. Adapting image generation models for specific purposes

Although large generative models are trained on datasets with millions of images, fine-tuning them for specific tasks often requires far fewer samples. For example, models like Stable Diffusion can be adapted for a specific purpose using a method called Low-Rank Adaptation (LoRA) with as little as 20 image samples (Holostrawberry, 2025). LoRA significantly reduces the required computational cost and time, making fine-tuning of models accessible to users with consumer-grade GPUs (Martineau, 2024).

Originally developed for adapting LLMs to specific tasks such as analysing legal documents, LoRA is now also widely used in image generation. While previous techniques required retraining all model weights to fine-tune the model, this method focuses only on a subset consisting of the model's attention layers while freezing the rest. The targeted layers are responsible for ensuring that the generated images match the text prompts (Hu et al., 2022).

Numerous platforms on the Internet, such as CivitAI, host user-made LoRA models with a wide variety of model customisation targets, such as the representation of popular fictional characters and celebrities, different poses, facial expressions, objects, clothing and backgrounds. There are also many models that aim to imitate the artistic style of contemporary artists, often without their consent.

On CivitAI, for example, a user has published a LoRA model that imitates illustrations by American comic artist Evan Stanley from IDW Publishing. The developer of this LoRA model has commented that commissioned work from an artist is too expensive, so they have trained this model to give others a free alternative. The model was used to generate over 62 thousand images and has been downloaded almost a thousand times. The model description does not indicate whether permission to use Evan Stanley's work in training was obtained from her, but the nature of the model suggests that this was probably not the case (AcanthAI, 2024).

2.4. Image generation practices and artist involvement on social media platforms

With the rise of generative AI, several social media platforms have begun training their own AI models on their users' data, including images. Many have updated their terms of service to allow such use. Often it is enabled by default and opting out is difficult or even impossible. Instagram, for example, with over two billion registered user accounts (as of February 2025), trains Meta AI services with its users' data. Opting out requires finding and filling out a hidden form in the app settings, in which you have to provide a reason for opting out. It only applies to future posts and is not available in many regions, including the United States (Jiménez, 2024).

Social media platform X (formerly Twitter) introduced similar rules in November 2024 to train its AI tool Grok (Pauley, 2024). So did Pinterest, a platform that was often used by artists as a source of inspiration, but which they are now abandoning due to the flood of AI-generated images (Dupré, 2025). Pinterest trains its image generator, Pinterest Canvas, with user data by default, regardless of the date the image was published (Pinterest, 2025).

In contrast, some platforms have adopted a more ethical stance. Adobe's artist platform Behance, for example, states that no user data is used to train its AI tool Adobe Firefly. They explain that it is only trained with public domain samples, as well as Adobe Stock data, with compensation paid to the authors (Adobe, a).

Nowadays, artists often rely on social media to promote their artwork and find employment opportunities. This is especially for emerging artists. For example, Latvian illustrator Paula Bobrova was discovered through her Instagram profile and was subsequently hired for the animated film "Flow", which later won the Golden Globe and Oscar awards (Dumbere, 2024; NFC, 2025a; NFC, 2025b). Bobrova created sketches of animal characters and the film's logo. This example illustrates why many artists, especially digital illustrators, cannot afford not to publish their work online, even if it sounds like it is the only viable way to avoid their artwork being used in AI training without permission.

Overall, given the uncertainties in the legislation, the misleading policies of social media platforms, and the ease of fine-tuning image generation models, we conclude that publicly available works by any artist are at risk of being incorporated into AI model training datasets. Therefore, artists need to protect their images and thus secure their copyright. The following chapter looks at the solutions currently available to protect images for this purpose.

3. Image protection software overview

This chapter deals with software solutions for the protection of artworks on the Internet. Three protection software – Glaze, Nightshade and Mist – are first described and then compared. The functionality and technical requirements of the individual software are given so that their practical suitability for artists without technical knowledge and/or necessary computer resources can be assessed. We also show how the images processed with the software look with different protection intensity settings.

3.1. Overview and comparison of Glaze, Nightshade and Mist

Glaze is an image protection software developed in 2023 by the Department of Computer Science at the University of Chicago to combat the unauthorised use of artwork when training and fine-tuning image generation tools to mimic artistic styles. The tool provides protection by masking the artist's work with a sufficiently different artistic style, chosen from a set of public domain images. This is achieved by using a pre-trained style transfer model and adding the resulting generated image to the original in the form of barely perceptible adversarial perturbations (UChicago, a).

If an image generation model is trained on multiple "glazed" images, it begins to associate the artist with the incorrect artistic style, and the resulting AI-generated images fail to successfully imitate the artist's original works. The Glaze team's study suggests that sufficient protection can be achieved if only 25% of an artist's online portfolio is "glazed" (Shan et al., 2023a).

Figure 1 illustrates the available perturbation intensities of Glaze. The difference from the original image is obtained by superimposing the processed image on the original and performing a contrast correction operation to improve visibility. The darker the pixels, the fewer differences there are with the original, while the lighter and brighter pixels indicate stronger perturbations, which become visibly discernible at higher settings. The following images for Figure 2 and Figure 3 were created in a similar way.

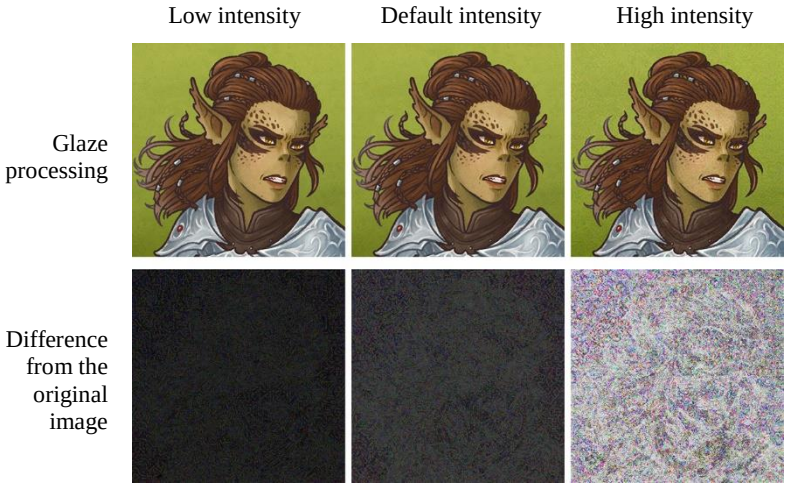

Figure 1. Sample processing with different Glaze perturbation intensities.

However, Glaze does not guarantee complete security for artworks. There have already been attempts to test the resilience of imperceptible perturbations against removal. Already in 2023, a group of professors and graduate students from Pennsylvania State University, Stony Brook University and the University of Illinois, Urbana Champaign in the USA proposed a "purification" method called IMPRESS, attempting to evaluate several contemporary protection methods, including Glaze (Cao et al., 2023). While this method demonstrated partial success on historical artistic styles, the developers of Glaze argue that it fails on contemporary works – its primary protection target (Shan et al., 2023b). Another study suggested several types of image manipulation that are effective in removing Glaze's changes (Hönig et al., 2024). Subsequent release of Glaze version 2.1 improved the robustness of the software against the "purification" methods (UChicago, 2024).

Nightshade, also developed by the University of Chicago's Department of Computer Science in 2023, is designed as an offensive tool that turns images into "poisoned" samples that disrupt and degrade the performance of image generation models by exploiting

"concept sparsity" – the relatively limited representation of specific concepts (e.g. "cat", "forest", "impressionism") in the training datasets of large-scale models (UChicago, b).

Nightshade improves on primitive types of data poisoning attacks by subtly manipulating the features of the image at the pixel level while preserving the description that matches the visual content. For example, an image showing a dog is correctly described as a "photo of a dog" but contains small perturbations that change the representation of its features closer to the appearance of a cat in the eyes of the AI model. A successful Nightshade attack, which changes the model's understanding of a concept to an incorrect one, is achievable even with 50 "poisoned" samples and it has an additional effect on related concepts. For instance, "poisoning" the concept of "dog" affects the model's perception of "puppy", "husky" or "wolf", teaching it to generate creatures that look closer to cats (Shan et al., 2024).

Figure 2 illustrates Nightshade version 1.0.2 processing with different intensity levels. As can be seen, the low and default intensity images are visually similar to each other, but the high intensity image results in noticeable artefacts.

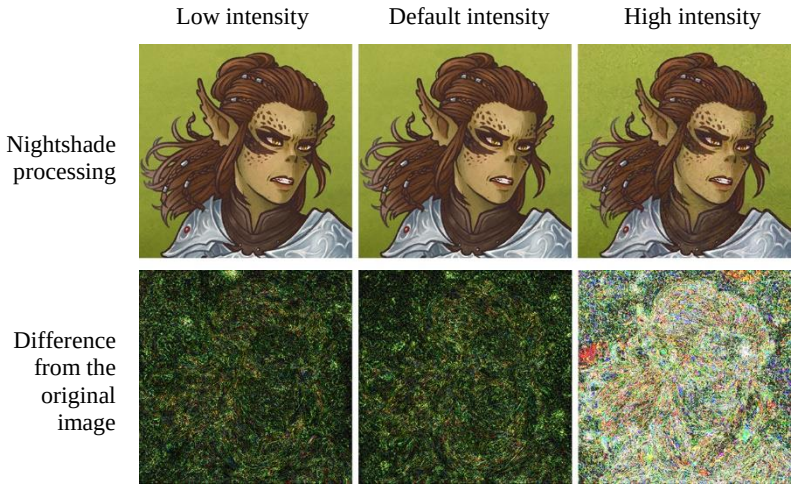

Figure 2. Nightshade "poisoned" samples with different levels of intensity.

Mist, developed in 2023 by the Psyker Group, is a targeted attack aimed at model classifiers, in contrast to Glaze and Nightshade, which aim to mislead models without degrading output quality. It works by embedding a specially selected chaotic pattern into images that causes the AI models to generate lower quality images, rendering the output unusable.

Mist uses two specially designed patterns that have high black and white contrast, frequent repetition and similarity to Moiré patterns. One incorporates the Mist logo, while the other – the NeurIPS logo. An image generation model trained on Mist-processed images begins to reproduce the chaotic patterns, thereby degrading the overall visual quality of all generated images (Zheng et al., 2023).

Figure 3 demonstrates images processed with different Mist version 2.0 intensity settings, ranging from 0 to 32. The default intensity is 12, while 1 was chosen as low

intensity and 32 as high. Compared to Glaze and Nightshade, Mist perturbations are significantly more visible. This is especially true for the highest intensity, where the used NeurIPS logo is easily perceptible.

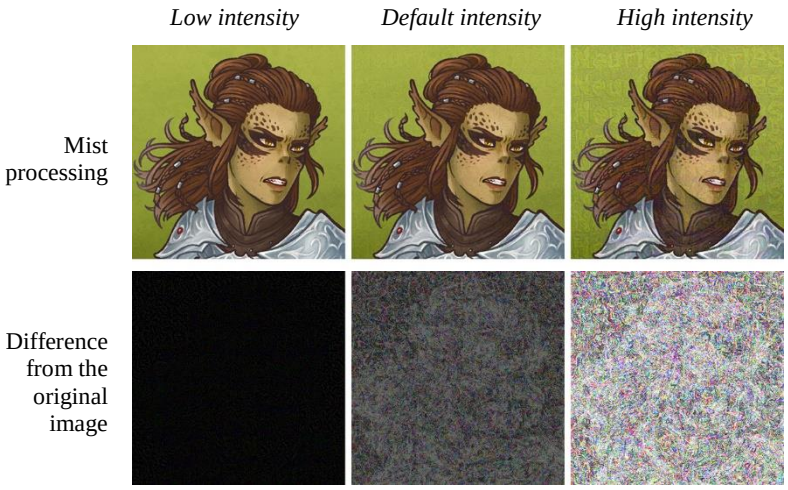

Figure 3. Mist processing samples with different levels of intensity.

Table 1 provides an overview of the three image protection tools and outlines the advantages and limitations of each using five criteria: Approach (image protection technique), Availability (required operating systems and/or existence of an online version), Requirements (minimum video random-access memory), Advantages, and Limitations. Furthermore, it should also be noted that all three tools share a drawback – they are hardware-intensive, which may be unsuitable for artists whose personal computers lack the necessary GPUs, or for those who do not use a computer to create art (for example, some use only iPads or traditional materials).

Table 1. Comparison of artwork protection tools Glaze, Nightshade and Mist

Criteria	*Glaze*	*Nightshade*	*Mist*
Approach	Facilitating incorrect style learning by masking the image with public domain artists' art styles	"Poisoning" models by exploiting concept sparsity	Introducing chaotic patterns into images to degrade the output's quality
Availability	• Windows/macOS • Web version WebGlaze	• Windows/macOS	• Windows/Linux • Google Colab notebook
Requirements	5 GB VRAM	5 GB VRAM	6 GB VRAM
Advantages	• Least visible perturbations • Easy to use, accessible (in terms of platform) • Receives the most updates	• Easy to use • Effective with relatively few poisoned samples • Affects related concepts	• Accessible (in terms of platform) • Approach theoretically provides equal protection for all art styles
Limitations	• Difficult to get a WebGlaze invite • Effectiveness depends on the user's art style	• Lack of an online version • Possibly unsuitable for illustrations depicting more than one clear subject	• Cannot run CPU mode on devices with non-NVIDIA GPUs • Errors and lack of user-friendly customisation in the Colab version • Most visible perturbations

3.2. Other types of protection software

Apart from these adversarial perturbation application tools, there are several other approaches to protect images from unauthorised use in the training of image generation models.

ArtShield applies an invisible watermark to protect images from being automatically scraped for training datasets. It mimics the watermarks used by AI image generation models to prevent AI-generated images from entering training datasets. The watermark is embedded by converting the image from RGB to YUV channels and applying discrete wavelet transforms (Xie, 2023). However, ArtShield does not protect artwork from users who can manually download any image from the Internet to train AI models with it.

Nepenthes and Iocaine are anti-scraping methods that function as "digital tarpits". They trap web crawlers that do not respect "robots.txt" anti-crawl directives in an endless maze where they are fed incomprehensible data created by a Markov babbler that produces text mimicking the structure of English sentences but lacking any meaning (Belanger, 2025). Although these methods were developed to protect websites from text scraping, they could potentially be repurposed by artists using personal portfolio websites.

Cara and ArtGram are social media platforms that were created for publishing artworks while disallowing the posting of AI-generated images. Cara is integrated with Glaze, which registered users can use instantly to protect their published works, while ArtGram claims to protect users' works with unique identifying signatures. In addition, Cara offers the possibility of posting job opportunities for artists, and ArtGram offers an online store with materials for artists and other creatives.

Finally, Have I Been Trained (https://haveibeentrained.com/) is a website where users can check whether their artworks appear in publicly available datasets. Using this website, a user can request to exclude their works from future model training, but, of course, there is no way to retroactively remove them from already trained models.

It is clear from this overview of adversarial perturbation applications Glaze, Nightshade and Mist that their use may in many cases be difficult for artists who (a) are not specialists in the field of computer science and (b) do not have access to sufficiently powerful computer equipment. On the other hand, other types of protection software can only partially protect images. Currently, there is a lack of practical comparative experimentation with the three tools, which we aim to provide further in this article. It is also important to determine which of these methods artists find the most effective and visually acceptable.

In order to assess how exactly artists and artwork viewers (non-artists) evaluate the protection of images, the authors trained several small AI models and then developed a survey. This is discussed in the next chapter.

4. Research Design

Here we describe the preparation of our custom dataset, the fine-tuning of the Stable Diffusion model and the process of sample generation with the trained LoRA models. We conclude this chapter by describing how the survey was designed and conducted. Finally, we describe the computational metrics used to compare with survey results.

4.1. Dataset preparation

To ensure the authenticity of the results, it was decided to create an independent dataset instead of using publicly available collections. This was further motivated by findings of

the Glaze team that using public domain artworks might not be as effective in testing, as these works could already be present in the training datasets of large-scale models. Additionally, this choice provided an opportunity to evaluate the protection performance, particularly for digital illustrations, which are more difficult to protect than, for example, online reproductions of paintings executed initially on canvas in oil or other material techniques.

As established in Section 2.3, LoRA models adjust only a small selection of model weights and require around 20 or more training samples to achieve style imitation results (Holostrawberry, 2025). Therefore, for our dataset, we chose 20 digital drawings created by one of the authors of the article within the last four years using the digital art software Paint Tool Sai and the graphic tablets Wacom Intuos Pro Medium and Huion Kamvas Pro 16. All drawings depict stylised portraits of various characters on a coloured background rendered in a consistent digital technique. These particular drawings were never published online, which ensures their absence from any online training datasets. The drawings were cropped and saved as .png files with a resolution 512 x 512 pixels.

A description was created for each drawing and stored in separate .txt files. These descriptions were initially generated using the open source image description model Large Language and Vision Assistant (LLaVA) and manually corrected to improve accuracy (Hugging Face, 2023). The descriptions included distinguishing features of the depicted figures such as gender, age, hairstyle, hair and eye colour, clothing, accessories, facial expression, pose and visible additional objects. The files of the drawings and descriptions were numbered uniformly in pairs (e.g. 1.png and 1.txt), and the prepared dataset was uploaded to Google Drive for further processing. Figure 4 presents selected sample images from the prepared dataset.

Female, silver hair, bob, grey eyes, white shirt, white bow tie, smirking, beauty mark, blue earrings	Male, short hair, brown hair, serious, brown eyes, white shirt, red vest, blue tie, gold bracelet, raised fist	Female, green hair, fluffy hair, green eyes, black and white dress, striped gloves, surprised, slightly open mouth, hand in hair, small robot on shoulder	Male, medium hair, brown hair, brown beard, brown eyes, smiling with teeth, purple jacket, silver earring

Figure 4. Dataset samples. The activation word “valstyle” and the keyword “character” identified by Nightshade were also included in each prompt. Author of the images – Valerija Januseva.

Next, the dataset was processed with the selected protection software. In total, eight datasets were created for the experiment:

- Original images without protection,
- Glaze protection (default and high intensity),
- Nightshade protection,
- Nightshade and Glaze combination,
- Mist protection (low, default and maximum intensity).

The processing with Glaze and Nightshade was done using the respective desktop software. Due to the lack of powerful computer resources, running the software in CPU mode took quite a long time – processing an image with Glaze took two to three hours, and with Nightshade – one hour. The Mist desktop software had inexplicable runtime failures that could not be resolved. Therefore, the online Google Colab notebook was used instead. It took about two minutes and 34 seconds to process one image, and the entire dataset was completed in 51 minutes and 21 seconds. Using Google Colab required resolving some errors by modifying the requirements.txt file.

4.2. Training image generation models

For privacy and security reasons, model training was conducted using entirely open source resources and a private workspace. The widely used text-to-image model Stable Diffusion 1.5 was used as the base model. The Python code for fine-tuning was deployed in the format of a Jupyter Notebook on Google Colab, utilising the platform's free NVIDIA Tesla T4 GPU resources with 16 GB of VRAM. To access Stable Diffusion, a free Hugging Face account was created. The authentication token obtained was set as a secret parameter (HF_TOKEN) in the Google Colab environment. For fine-tuning, the LoRA method was chosen due to its computational efficiency and suitability for the style imitation task. Initially, an attempt was made to use LoRA training scripts made by GitHub user *kohya_ss* (Kohya, 2022). As it caused some runtime failures, a more Colab-friendly modification by GitHub user *hollowstrawberry* was implemented instead (Hollowstrawberry, 2023).

The hyperparameters were adjusted to optimise the quality of image generation. Table 2 summarises the final hyperparameter configuration used across all LoRA models. The selection of values was based on a combination of recommendations for LoRA training (Hollowstrawberry, 2023) and the authors' own experimentation. For example, a learning rate that is too low ($1\times10^{-4)}$ results in the generation of overly realistic faces that do not match the original works' stylisation, while a rate that is too high (1×10^{-3}) produces broken, chaotic output. It is also important to balance other hyperparameters accordingly. Therefore, we set the U-Net learning rate to 5×10^{-4}, while the text encoder learning rate was kept lower at 1×10^{-4} to balance visual and textual learning. We set repeats to 20, batch size to 2, and trained over 10 epochs, resulting in a total of 2,000 steps. On average, the training of one LoRA model required approximately 32 minutes.

Table 2. Hyperparameter configuration for model training

Hyperparameter	*Explanation*	*Value*
U-Net learning rate	Controls how much the model weights are updated with each step to learn visual elements and structure.	5×10^{-4}
Text encoder learning rate	Controls how much the model weights are updated with each step to learn textual descriptions. It is recommended to use a lower value than the U-Net learning rate.	1×10^{-4}
Repeats	Dataset image repetitions during training.	20
Batch size	The amount of training data in each training round.	2
Epochs	The number of iterations during training of the entire dataset fed to the model.	10
Steps	The total number of iterations the model processes each batch of data to update the weights. Total steps = epochs * (dataset size * repeats / batch size).	2,000

4.3. Sample generation

A separate LoRA model was fine-tuned with each of the dataset variants mentioned in Chapter 4.1. Multiple samples were generated by loading each model's weights as a .safetensors file into the base Stable Diffusion 1.5 model using the AutoPipelineForText2Image pipeline from the Hugging Face Diffusers library.

The images were created in the same Google Colab environment as before. Loading the base model took about a minute, while loading the LoRA weights into the base model took about two seconds. To try another LoRA model in the same session, the base model was refreshed, which took another 20 seconds each time.

Five prompts similar to the description format used in the training dataset were created for image generation. They contained the activation word "valstyle", as well as a list of character features chosen in such a way that their combination was distinct from the original dataset samples. For each model, ten images were generated per prompt, and one of them was randomly selected for use in the survey.

Figure 5 shows samples generated with the trained LoRA models. Each row corresponds to a text prompt used in generating the corresponding image. The columns

represent the LoRA models that were trained on the specified dataset and used for generating the respective images.

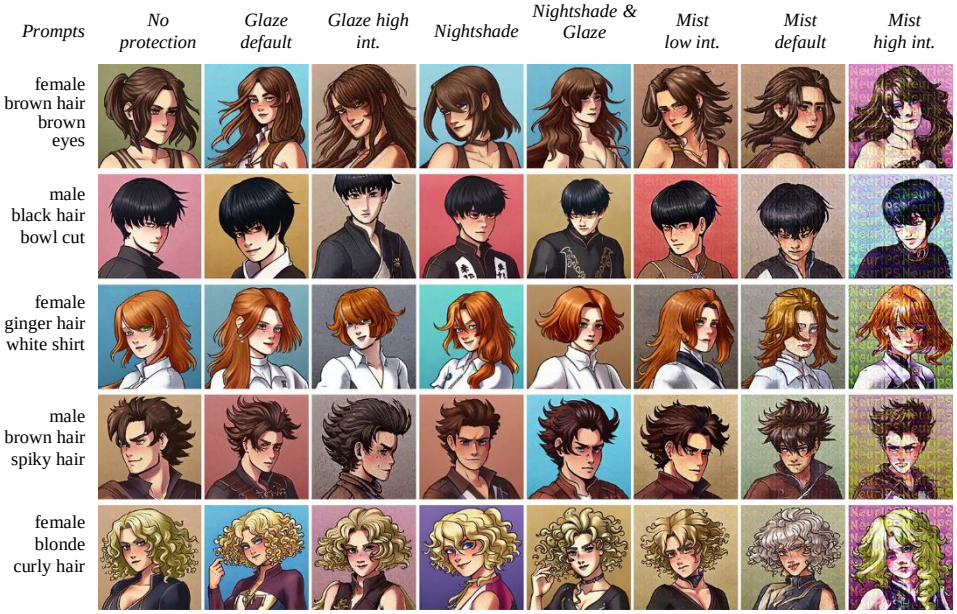

Figure 5. Sample images generated with the fine-tuned LoRA models. Each prompt also contained the activation word "valstyle" and the keyword "character".

4.4. Survey design

An anonymous user survey was conducted to evaluate the generated samples in terms of (a) similarity to original artworks, (b) overall image quality, (c) practical applicability and visual acceptability, and (d) agreement or disagreement with the use of image protection tools. To ensure that the survey questions met the above criteria for image analysis, a pilot test was conducted with a small group of selected participants from target groups of artists and non-artists. Based on the results of this pilot test, the wording of some questions was clarified, thereby improving the quality of the survey. The survey consisted of 41 questions and was divided into four parts.

1. The first section collected demographic data and also respondents' background, including involvement with creating art, to analyse whether there were differences between artists' and artwork viewers' (referred to as non-artists in the survey and in the description of the survey results) evaluations of artworks. It also included questions about respondents' experience with image generation and their views on the use of generative AI.

2. The second section was available only to artists and offered six questions asking about (a) the forms of art they work with, (b) the use of social media for publishing art and (c) awareness of protection software.

3. In the third section, respondents were shown both authentic drawings and style-mimicking AI-generated images. Respondents were asked to evaluate how close they think the generated samples were to the originals. They were told to consider (a) the stylisation of facial features, (b) the choice of colours, (c) the textures and (d) the overall image quality. The evaluation was based on a Likert scale.

4. In the fourth section, respondents were asked to rate the image quality as well as the practical applicability and visual acceptability of the illustrations after processing with protection software.

A total of 71 people took part in the survey, most of whom (99%) were between 18 and 36 years old. The participants included 45 artists as well as 26 artwork viewers (non-artists). The study was distributed among University of Latvia students and (professional and hobbyist) artists. These focus groups were selected to elicit the opinions of artists and non-artists. A deviation in group distribution occurred because some students were also hobbyist artists. The authors of the paper accepted this shift, believing it reflects the contemporary situation in which, thanks to the democratisation of digital art creation tools, representatives of other professions are actively working as hobbyist artists. Moreover, it coincided with the empirical observations of one of the authors, who is herself engaged in digital illustration as a hobby.

60% of participating artists worked both traditionally (drawing, painting, etc.) and digitally (digital illustration, graphic design, 3D modelling, etc.). Six artists worked only with traditional and twelve only with digital art creation techniques. One artist additionally specified their work with traditional printmaking techniques – lithography, letterpress and serigraphy.

Finally, to complement the survey with computational metrics, we have extracted features from images by using the VGG-19 convolutional network, as well as computed image embeddings using the open source openCLIP model (Simonyan and Zisserman, 2014; MLfoundations, a). For both, we have calculated the average cosine similarity between sets of images and expressed it in percentages. We used the original 20 artwork dataset as the reference to which compare each of the generated datasets, both with and without protection.

The next chapter analyses the results of the study and highlights the strengths and areas for future improvement of screen protection tools from the respondents' perspective.

5. Results

The results of the survey are presented in six figures and three tables. They show the respondents' attitudes towards image generation as well as their assessment of various aspects of the quality of protection offered by the generated images.

Figure 6 shows the frequency of use by respondent groups when asked how often they use AI image generation tools. Most respondents had a negative attitude towards image generation with AI tools. 21.1% were neutral, and only two respondents had a positive opinion. Despite the observed negative attitude, more than half of the respondents have

used image generation tools at least occasionally. 68.4% of respondents who have tried using AI tools at least a couple of times have used them strictly for entertainment, while 13.2% have utilised them for professional purposes. Several artists indicated that they were forced to use AI image generation tools for assignments at university or school. Some other artists used these tools to gain inspiration, as well as for quick visualisation of ideas and concepts required by their workplace. One non-artist had considered using image generation for text visualisation for children, but in the end decided not to use such tools. In addition, when respondents were asked about their habits when using AI, it was also found that 74.6% of them had observed cases where someone had tried to use artificial intelligence to imitate the artistic style of a contemporary artist they knew. Two artists stated that someone had specifically tried to imitate their artistic style.

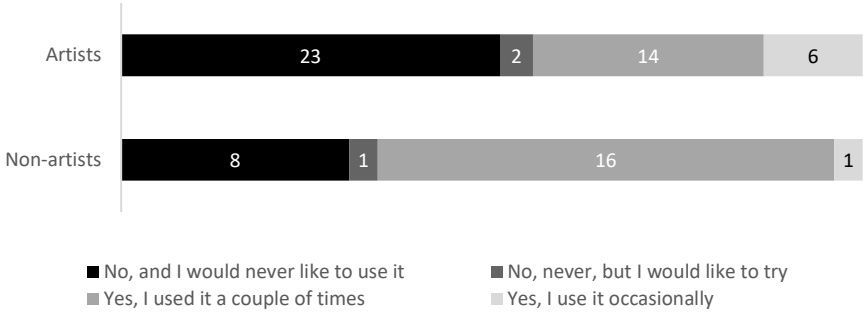

Figure 6. Frequency of use of image generation tools in the artist and non-artist groups.

80% of the surveyed artists publish images of their work online. As can be seen in Figure 7, the most popular social networking platforms among them are Instagram, X (formerly Twitter), and Tumblr. One artist uses a personal portfolio website.

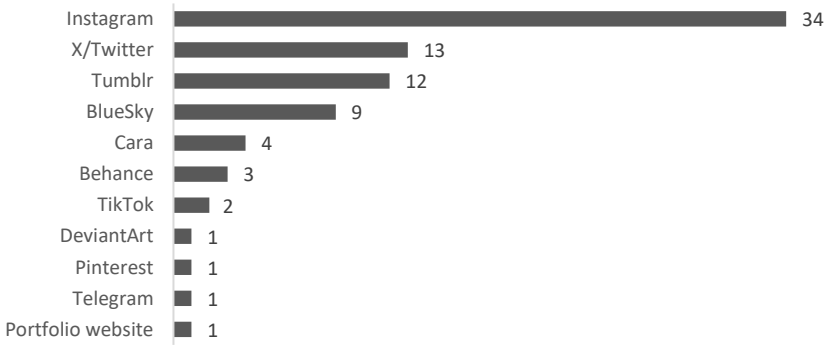

Figure 7. The choice of artists for publishing images on social media.

66.67% of artists do not protect their work when publishing it online. Those who do usually only add a watermark or signature. Only four artists use Glaze, two use Nightshade, and none use Mist. Four artists use the Cara platform, which integrates Glaze. The artist who created a personal portfolio website stated that their copyright is described there.

Figure 8 shows the distribution of average similarity ratings, divided into four individual criteria – facial feature stylisation, colour choices, textures and overall quality of the image. The lower the percentage rating, the worse the results of the style imitation. As can be seen, the models do not mimic the texture and image quality of the original illustrations as well as the colours and facial feature stylisation. Particularly poor texture and quality are observed for samples with high Glaze intensity and all Mist samples.

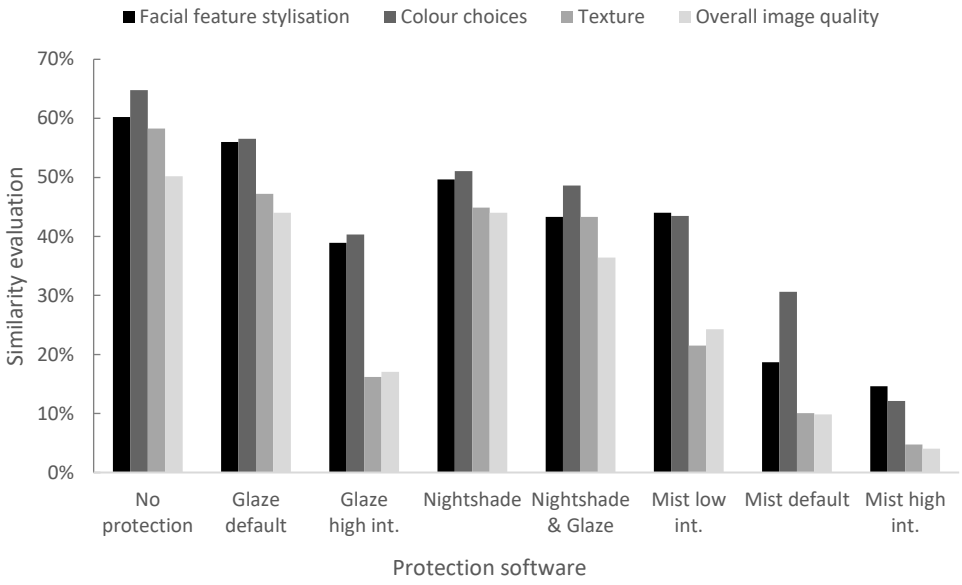

Figure 8. Evaluations of similarity to original illustrations.

Figure 9 shows the combined average similarity ratings of all respondents, divided into artist and non-artist groups. The higher the percentage, the greater the similarity to the original illustrations. The closest to the originals were the samples that were generated without any protection. However, their ratings were lower than expected – 56.18% (artists) and 62.13% (non-artists). As can be seen, the ratings for all Mist intensities and high Glaze intensity are lower compared to other software. The ratings for default Glaze, Nightshade and the combination of Nightshade and Glaze are closer to the generated images without protection, so the effectiveness of these settings is not as strong.

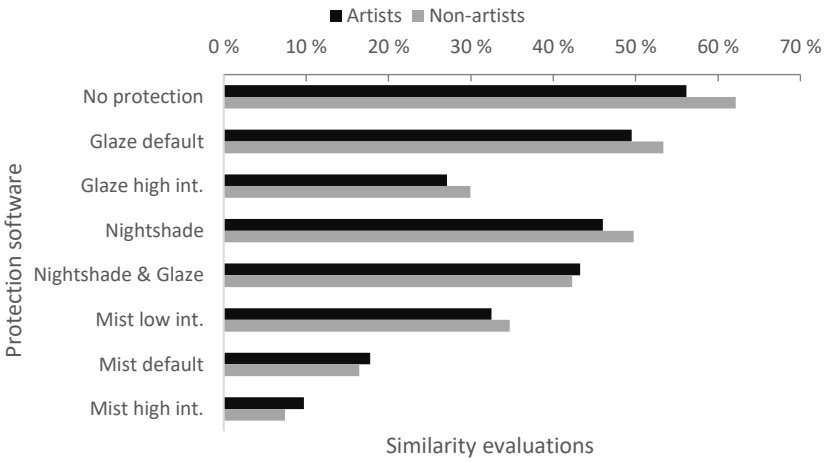

Figure 9. Assessments of the similarity of the generated images to the original illustrations.

Figure 10 shows the rating of the quality and practical applicability of protected images by group. The higher the percentage, the more respondents agreed to apply the respective perturbations to the images to achieve the specified level of protection. The low and default intensity Mist samples have the highest ratings – 80% and 82.22% of artists were satisfied with the visual intensity of Mist perturbations and the offered level of protection. Ratings of Glaze, Nightshade and their combination are generally lower, with the exception of high intensity Glaze protection. In almost all cases, artists were more willing to use higher intensity perturbations than non-artists.

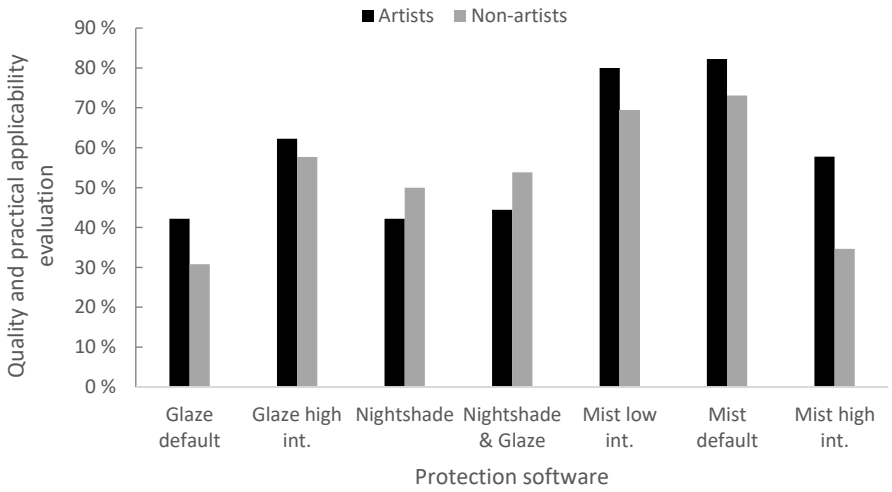

Figure 10. Assessment of the quality and practical applicability of protected images.

Figure 11 reveals the respondents' explanations as to why they would not agree to apply perturbations with protection software, divided into artist and non-artist groups. High intensity Mist perturbations are the only case where both artists and non-artists would reject the application because of artefacts of processing being very visible. Meanwhile, Glaze, Nightshade, and their combination do not provide sufficient protection, especially according to artists.

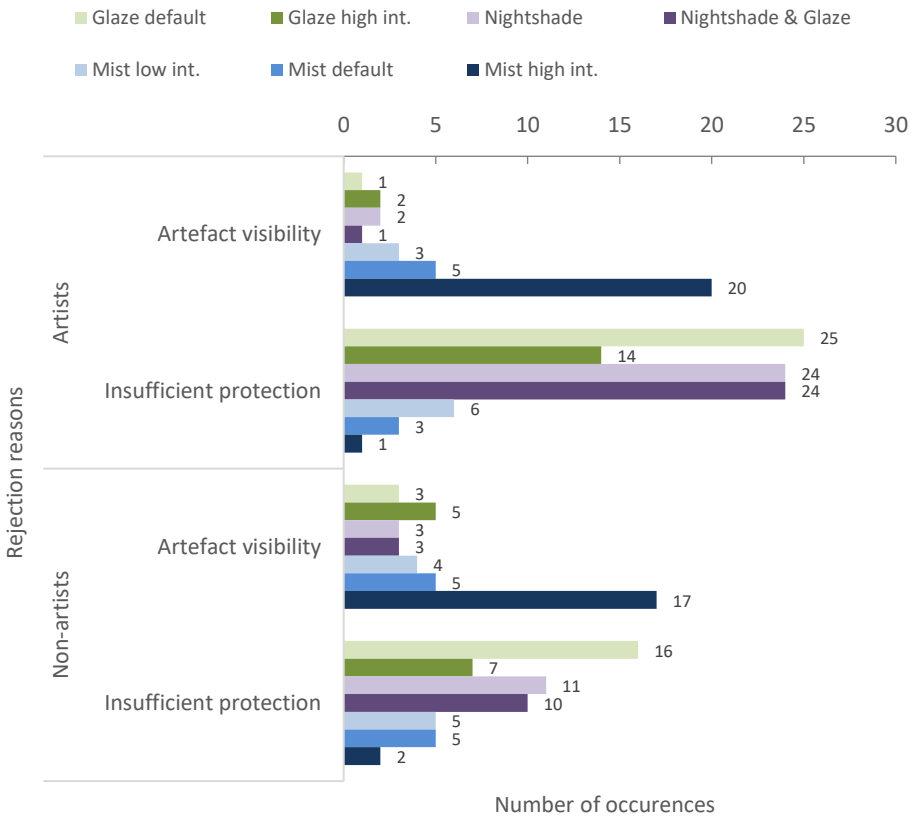

Figure 11. Reasons for rejecting the use of image protection software from the perspective of artists and non-artists.

Table 3 summarises the survey results of the evaluation of the images. The higher the percentage value, the more successful the imitation of the original style. Since the generated images without protection only received a similarity score of 58.36%, the samples whose evaluations were at least 20% below this value were considered successful protection.

Table 3. Evaluation results of the similarity of the generated images with the originals. Lower percentages here indicate more effective protection by the tools – such results are marked in green

Protection	*Average evaluation*	*Facial features*	*Colour choices*	*Texture*	*Image quality*
None	58.36%	60.21%	64.79%	58.27%	50.18%
Glaze default	50.92%	55.99%	56.51%	47.18%	44.01%
Glaze high int.	28.13%	38.91%	40.32%	16.20%	17.08%
Nightshade	47.40%	49.65%	51.06%	44.89%	44.01%
Nightshade & Glaze	42.91%	43.31%	48.59%	43.31%	36.44%
Mist low int.	33.32%	44.01%	43.49%	21.48%	24.30%
Mist default	17.30%	18.66%	30.63%	10.04%	9.86%
Mist high int.	8.89%	14.61%	12.15%	4.75%	4.05%

Table 4 reveals the assessment of image quality and visual acceptability ratings of the protected images. Ratings above 50% are considered acceptable, i.e. most respondents would agree to the application of the appropriate protection to the artwork.

Table 4. Results of the evaluation of protected images. Ratings above 50% are marked in green

Protection/criteria	*Agree to use*	*Disagree to use*	*Most common disagreement reason*
Glaze default	38.03%	61.97%	Insufficient protection
Glaze high int.	60.56%	39.44%	Insufficient protection
Nightshade	45.07%	54.93%	Insufficient protection
Nightshade & Glaze	47.89%	52.11%	Insufficient protection
Mist low int.	76.06%	23.94%	Insufficient protection
Mist default	78,87%	21,13%	Artefact visibility
Mist high int.	49,30%	50,70%	Artefact visibility

Table 5 shows both the survey results and the results obtained by computing style similarity. The higher the evaluation, the closer that dataset is to the original artworks.

Table 5. Overview of the evaluation of unprotected and protected images

Protection	*Survey evaluation*	*VGG-based evaluation*	*CLIP-based evaluation*	*Agree to use*
None	58.36%	97.27%	95.77%	-
Glaze default	50.92%	96.17%	95.43%	38.03%
Glaze high int.	28.13%	94.51%	93.87%	60.56%
Nightshade	47.40%	97.72%	96.04%	45.07%
Nightshade & Glaze	42.91%	95.86%	95.56%	47.89%
Mist low int.	33.32%	95.16%	92.80%	76.06%
Mist default	17.30%	88.76%	88.87%	78.87%
Mist high int.	8.89%	73.64%	78.01%	49.30%

Both the VGG-based and CLIP-based similarity measures showed high stylistic similarity across all protection software and settings (ranging from approximately 73% to 97%). These scores are noticeably different from the subjective survey evaluations. This discrepancy suggests that the neural models capture some visual or structural features that do not fully correspond to human perceptions of artistic style or visual image quality. Therefore, the objective similarity scores should be interpreted as indicators of representational closeness in feature space, rather than how they appear to human viewers. Mist received lower similarity scores across all metrics, both model-based and survey evaluations, while Glaze and Nightshade are of higher similarity to the original illustrations.

Other studies (Shan et al., 2023a) indicate that Glaze has higher effectiveness for styles that are closer to traditional paintings, but is currently limited for simpler illustration

styles. It corresponds with the low ratings of Glaze in our evaluations – at least with the default intensity settings. The combination of Nightshade and Glaze leads to stronger protection against style imitation, but is still considered insufficient by almost half of the survey respondents.

The Nightshade protection results did not demonstrate concept "poisoning", most likely because the experiment conducted in this study was different from the experiments conducted by the Nightshade developers. However, the perturbations still affected the model training and slightly impaired the model's ability to imitate the style of the illustrations.

Mist received the highest scores for style imitation protection, image quality and practical applicability. Default intensity was preferred. 78.87% of respondents would agree to use it, and the overall similarity of the generated images to the original illustrations was rated at 17.30%. While the similarity rating for high intensity Mist is even lower (8.89%), only 49.30% of respondents would want to use images with such highly visible perturbations. Although Glaze provides a protection method that could better prevent "purification" attempts, it does not offer the same level of protection for all illustration styles. In contrast, the targeted perturbations offered by Mist protect all image types equally, which may be more appealing to artists.

Research results show that the hypothesis set at the beginning has been partially confirmed. Of the tools analysed, Mist affects the models' ability to generate images the most – it significantly reduces the quality of the generated outputs. On the other hand, the results of Glaze and Nightshade, as well as their combination, are too close to those of the unprotected artworks. Therefore, it is difficult to consider them completely successful, apart from using the high intensity settings of Glaze. However, despite the shortcomings, the majority of respondents acknowledge that when publishing works of art online, they need to be protected.

6. Discussion

This article contributes to the emerging field of digital artwork protection by providing a practical comparison of three image protection tools – Glaze, Nightshade, and Mist – which aim to protect artists from unauthorised imitation of their artworks by image generation models.

Although the developers of each software have conducted their own experiments, these used separate experiment designs and training data, making it difficult to draw precise comparisons between these three tools. Unlike their studies, our experiments were designed to compare the tools using a unified methodology. We evaluated both technical effectiveness (similarity between original and generated images) and visual usability (acceptability of the intensity of perturbations) of these tools. Furthermore, we tested these tools on digital illustrations and art, a type of artwork creation medium that has not been well researched.

Mist proved to be the most effective tool, reducing similarity to original artworks to 17.30% at default intensity while maintaining high acceptability (78.87%). In contrast, Glaze and Nightshade showed similarity scores above 42%, close to unprotected images (58.36%), and were rated as insufficient by most respondents (61,97% and 54,93%

respectively). The VGG-based and CLIP-based evaluations, although much higher in similarity, still correlate with the survey evaluations. These results suggest that protection methods like Glaze and Nightshade may not adequately protect digital illustrations, highlighting the benefits of Mist's targeted perturbations.

These findings indicate that image protection is not only theoretically possible but also supported by the artist community. Additionally, the trade-off between protection strength and visual acceptability was demonstrated: while high intensity Mist achieved the lowest similarity score (8.89%), fewer than half of respondents (49.30%) considered the outputs usable due to the visibility of artefacts. This highlights the importance of developing tools that balance technical effectiveness with visual quality.

We acknowledge that our study has limitations, primarily because our experiments were conducted on illustrations by a single artist. However, we have defined a methodology that could be useful for other researchers. Future work could continue our experiments with different artistic styles from several artists, which could be particularly useful for further evaluating the effectiveness of Glaze's style transfer.

Furthermore, lower results for models trained with unprotected samples indicate that the LoRA models used in the experiment could be improved to produce more accurate style imitation results relative to the original illustrations. The effectiveness of perturbation "purification" methods could also be further investigated.

It should be noted that the field of AI is still rapidly evolving, and the fight against unauthorised AI training could be a never-ending arms race. However, this research provides quantitative evidence of the advantages and limitations of existing image protection tools, and it establishes Mist as a potential candidate for further practical implementation. Our study also demonstrates that artists are now both aware of such technologies and are willing to use them. By combining technical experimentation with user-centred evaluation, this research strengthens the theoretical and practical foundation for protecting artworks in the era of generative AI.

7. Conclusions

This article investigated the effectiveness of image protection tools, such as Glaze, Nightshade, and Mist, against unauthorised imitation of artworks using AI image generation models. Using a combination of empirical research, technical experiments and a survey involving respondents including both artists and viewers of artworks (non-artists), this study offered insights into the current state of digital artwork protection. The results highlighted the need for protection tools, especially considering how easy it is to customise models with only 20-30 samples of artworks without the author's consent.

Of the tools tested, Mist demonstrated the most consistent performance, successfully deteriorating the quality of the model's output images even at low and default intensity settings, resulting in as low as 33.32% and 17.30% similarity to originals, respectively. Furthermore, Mist received lower similarity scores across all metrics, both model-based and survey assessments, whereas Glaze and Nightshade were more similar to the original illustrations. However, as the technical experiments have shown, the current version of the Mist software might be challenging to use for artists without technical knowledge. The tool needs to be improved to make it more user-friendly for everyone, regardless of

technical skills. However, 78.87% of the artists who participated in the survey were willing to use Mist's default intensity settings for image protection. This suggests that image protection is practically possible, and artists support it.

Overall, the findings confirm that practical image protection is feasible and supported by artists, though improvements in usability and balance of technical effectiveness and visual quality remain necessary. The study provides a theoretical and practical knowledge base for those interested in protecting their artwork and for further research.

Acknowledgements

We would like to express our sincere gratitude to Professor Juris Borzovs for suggesting the publication of this article and for his helpful advice on improving its contents. We are also grateful to the reviewers for their feedback and recommendations.

References

AcanthAI (2024). *Sonic IDW Style (Evan Stanley)*, available at https://civitai.com/models/596723/sonic-idw-style-evan-stanley.

Adobe (a). *Our approach to generative AI with Adobe Firefly*, available at https://www.adobe.com/ai/overview/firefly/gen-ai-approach.html#.

Belanger, A. (2025). *AI haters build tarpits to trap and trick AI scrapers that ignore robots.txt*, available at https://arstechnica.com/tech-policy/2025/01/ai-haters-build-tarpits-to-trap-and-trick-ai-scrapers-that-ignore-robots-txt/.

Bergmann, D., Stryker, C. (2024). *What is a variational autoencoder?*, available at https://www.ibm.com/think/topics/variational-autoencoder.

Brittain, B. (2023). *AI-generated art cannot receive copyrights, US court says*, available at https://www.reuters.com/legal/ai-generated-art-cannot-receive-copyrights-us-court-says-2023-08-21/.

Cao, B., Li, C., Wang, T., Jia, J., Li, B., Chen, J. (2023). *IMPRESS: Evaluating the Resilience of Imperceptible Perturbations Against Unauthorized Data Usage in Diffusion-Based Generative AI*, Advances in Neural Information Processing Systems, **36**, pp. 10657–10677.

Cho, W. (2024). *Artists Score Major Win in Copyright Case Against AI Art Generators*, available at https://www.hollywoodreporter.com/business/business-news/artists-score-major-win-copyright-case-against-ai-art-generators-1235973601/.

Coca-Cola (2024). *The Holiday Magic is coming*, available at https://www.youtube.com/watch?v=4RSTupbfGog.

Dumbere, L. (2024). *In the same boat* (in Latvian), available at https://ir.lv/2024/09/18/viena-laiva/.

Dupré, M. H. (2025). *Pinterest Changes User Terms So It Can Train AI on User Data and Photos, Regardless of When They Were Posted*, available at https://futurism.com/pinterest-data-photos-train-ai.

European Parliament (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council*, available at https://eur-lex.europa.eu/eli/reg/2024/1689/oj.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y. (2014). *Generative Adversarial Nets*, Advances in Neural Information Processing Systems, **27**.

Harding, S. (2025). *Netflix will show generative AI ads midway through streams in 2026*, available at https://arstechnica.com/gadgets/2025/05/netflix-will-show-generative-ai-ads-midway-through-streams-in-2026/.

Heikkilä, M. (2022). *This artist is dominating AI-generated art. And he's not happy about it*, available at https://www.technologyreview.com/2022/09/16/1059598/this-artist-is-dominating-ai-generated-art-and-hes-not-happy-about-it/.

Ho, K. K. (2024). *Database of 16,000 Artists Used to Train Midjourney AI, Including 6-Year-Old Child, Garners Criticism*, available at https://www.artnews.com/art-news/news/midjourney-ai-artists-database-1234691955/.

Hollowstrawberry (2023). *Kohya Colabs*, available at https://github.com/hollowstrawberry/kohya-colab.

Holostrawberry (2025). *Make your own Loras, easy and free, using Colab*, available at https://arcenciel.io/articles/1

Hu, E., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Chen, W. (2022). *LoRA: Low-Rank Adaptation of Large Language Models*, ICLR, **1**.

Hugging Face (2023). *LLaVa*, available at https://huggingface.co/docs/transformers/en/model_doc/llava

Hönig, R., Rando, J., Carlini, N., Tramèr, F. (2024). *Adversarial Perturbations Cannot Reliably Protect Artists From Generative AI*, available at https://arxiv.org/pdf/2406.12027.

Jiménez, J. (2024). *Worried About Meta Using Your Instagram to Train Its A.I.? Here's What to Know*, available at https://www.nytimes.com/article/meta-ai-scraping-policy.html.

Kohya, S. (2022). *sd-scripts*, available at https://github.com/kohya-ss/sd-scripts.

Martineau, K. (2024). *Serving customized AI models at scale with LoRA*, available at https://research.ibm.com/blog/LoRAs-explained.

McCormack, J., Cruz Gambardella, C., Rajcic, N., Krol, S.J., Llano, M.T., Yang, M. (2023). *Is Writing Prompts Really Making Art?*, International Conference on Computational Intelligence in Music, Sound, Art and Design (Part of EvoStar), pp. 196–211.

Milmo, D. (2024). *'Impossible' to create AI tools like ChatGPT without copyrighted material, OpenAI says*, available at https://www.theguardian.com/technology/2024/jan/08/ai-tools-chatgpt-copyrighted-material-openai/.

MLfoundations. (a) *OpenCLIP*, available at https://github.com/mlfoundations/open_clip.

MPNFW (2016). Manhattan Project for a Nuclear-Free World. *Hayao Miyazaki's thoughts on an artificial intelligence*, available at https://www.youtube.com/watch?v=ngZ0K3lWKRc.

Mullicane, E. D. (2025). *"A Violation of the Law": After ChatGPT's AI Attack on Studio Ghibli, Lawmakers Are Looking to Take Legal Action*, available at https://screenrant.com/studio-ghibli-ai-artwork-chatgpt-japan-lawmakers-illegal/.

NFC (2025a). National Film Centre. *"Flow" becomes the first Latvian film to win the "Golden Globes" award*, available at https://www.nkc.gov.lv/en/article/flow-becomes-first-latvian-film-win-golden-globes-award.

NFC (2025b). National Film Centre. *"Flow" brings Latvia first-ever "Oscar"*, available at https://www.nkc.gov.lv/en/article/flow-brings-latvia-first-ever-oscar.

OpenAI (2025). *Introducing 4o Image Generation*, available at https://openai.com/index/introducing-4o-image-generation/.

Parshall, A. (2023). *How This AI Image Won a Major Photography Competition*, available at https://www.scientificamerican.com/article/how-my-ai-image-won-a-major-photography-competition/.

Pauley, C. (2024). *Elon Musk's X Can Now Use Your Data to Train Its AI*, available at https://web.archive.org/web/20250328021044/https://9meters.com/entertainment/social-media/elon-musks-x-can-now-use-your-data-to-train-its-ai.

Pinterest (2025). *Privacy Policy*, available at https://policy.pinterest.com/en/privacy-policy.

Porterfield, C. (2023). *Judge dismisses most of artists' copyright lawsuit against AI image generators*, available at https://www.theartnewspaper.com/2023/10/31/california-judge-dismisses-most-of-artists-ai-copyright-lawsuit/.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B. (2022). *High-Resolution Image Synthesis with Latent Diffusion Models*, Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 10684–10695.

Roose, K. (2022). *An A.I.-Generated Picture Won an Art Prize. Artists Aren't Happy*, available at https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html/.

Shan, S., Cryan, J., Wenger, E., Zheng, H., Hanochka, R., Zhao, Y. B (2023a). *Glaze: Protecting Artists from Style Mimicry by Text-to-Image Models*, USENIX Security Symposium, **32**, pp. 2187–2204.

Shan, S., Wu, S., Zheng, H., Zhao, B. Y. (2023b). *A Response to Glaze Purification via IMPRESS*, available at https://arxiv.org/pdf/2312.07731.

Shan, S., Ding, W., Passananti, J., Wu, S., Zheng, H., Zhao, B. Y. (2024). *Nightshade: Prompt-Specific Poisoning Attacks on Text-to-Image Generative Models*, IEEE Symposium on Security and Privacy, pp. 807–825.

Shōyu (2025). *Shōyu Annual Birthday Party*, available at https://www.instagram.com/shoyu.riga/reel/DIbzdgvNdSB/.

Simonyan, K., Zisserman, A. (2014). *Very deep convolutional networks for large-scale image recognition*, available at https://arxiv.org/pdf/1409.1556.

Sohl-Dickstein, J., Weiss, E. A., Maheswaranathan, N., Ganguli, S. (2015). *Deep unsupervised learning using nonequilibrium thermodynamics*, International Conference on Machine Learning, **32**, pp. 2256–2265.

Sriram, A. (2025). *Ghibli effect: ChatGPT usage hits record after rollout of viral feature*, available at https://www.reuters.com/technology/artificial-intelligence/ghibli-effect-chatgpt-usage-hits-record-after-rollout-viral-feature-2025-04-01/.

UChicago (2024). University of Chicago. *A Note on the new attack paper: "Adversarial Perturbations Cannot Reliably Protect Artists From Generative AI"*, available at https://glaze.cs.uchicago.edu/update21.html.

UChicago (a). University of Chicago. *What is Glaze?*, available at https://glaze.cs.uchicago.edu/what-is-glaze.html.

UChicago (b). University of Chicago. *What is Nightshade?*, available at https://nightshade.cs.uchicago.edu/whatis.html.

WIPO (1979). *Berne Convention for the Protection of Literary and Artistic Works*, available at https://www.wipo.int/wipolex/en/text/283698.

Wyn Davies, C., Dennis, G. (2024). *Getty Images v Stability AI: the implications for UK copyright law and licensing*, available at https://www.pinsentmasons.com/out-law/analysis/getty-images-v-stability-ai-implications-copyright-law-licensing.

Xie, A. Z. (2023). *Stable Diffusion invisible watermarker - the math and our algorithm improvements*, available at https://artshield.io/blog/post/stable-diffusion-invisible-watermarker-the-math-and-our-algorithm-improvements.

Your Europe (2025). *Copyright*, available at https://europa.eu/youreurope/business/running-business/intellectual-property/copyright/index_en.htm#inline-nav-2.

Zheng, B., Liang, C., Wu, X. (2023). *Targeted Attack Improves Protection against Unauthorized Diffusion Customization*, available at https://arxiv.org/pdf/2310.04687.

Received August 20, 2025, revised October 15, 2025, accepted October 20, 2025