EVA Gallery Design Document

Digital Art Experience enhanced with AI and web3

@Teebor-Choka, @slobodaapl

August 22, 2024

Contents

Chapter 1

Features

1 AI: Art Theft Protection

Implementation of a multi-layered approach to safeguarding artists' work, including:

- **Nightshade:** A cutting-edge technique [1] that subtly alters artwork to render it useless for AI training while maintaining visual fidelity for human viewers.
- **NFT Minting (Non-AI pre-existing):** Using blockchain technology to create NFTs for each artwork, ensuring verifiable authenticity and ownership.
- **Metadata Integration:** Leveraging standardized metadata formats to store essential information about the artwork, including artist attribution, ownership details, and licensing terms.

1.1 Protection from use in AI models

Nightshade is a pioneering technique designed to protect artists' work from being exploited for training AI models. It introduces imperceptible perturbations to the image data, rendering it ineffective for AI training while preserving visual fidelity for human observers [1]. This approach ensures that the artwork remains visually appealing and unaltered for human consumption, while effectively "poisoning" the data for AI algorithms.

A similar tool, Glaze, focuses on protecting against style theft by subtly altering artwork to disrupt AI models' ability to mimic an artist's unique style [2]. However, Nightshade and Glaze are not yet interoperable, limiting their combined effectiveness in protecting against both content and style theft. EVA Gallery will monitor the development of these tools and integrate them when feasible to provide comprehensive protection against AI-powered art theft.

1.2 Plagiarism check: a two-pronged approach

EVA Gallery's plagiarism detection system employs a robust two-pronged approach to ensure the originality of uploaded artwork and protect artists' intellectual property rights, shown in Figure 1.1.

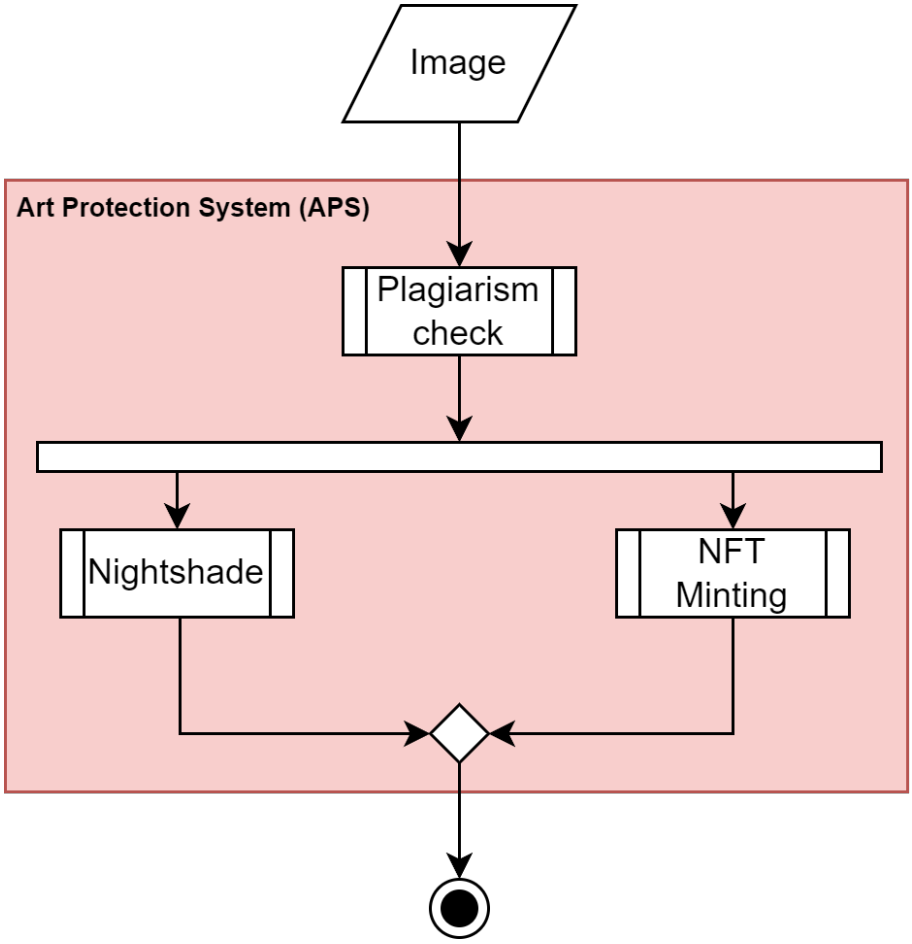

Figure 1.1: Data flow through the processes for the Art Protection System in EVA Gallery.

1.2.1 Embedding-based similarity check

The first step involves converting the uploaded artwork into a high-dimensional numerical representation called an embedding. This embedding captures the semantic content of the image, including its objects, scenes, and overall meaning. The embedding is then compared against all existing embeddings in the database using efficient similarity search techniques like HNSW indexing. If an extremely close match is found, the artwork is flagged for further inspection.

1.2.2 Gramian matrix similarity check

Flagged images undergo a second level of scrutiny using Gramian matrix analysis. A Gramian matrix is a mathematical representation that captures the style of an image, including its textures, colors, and patterns. It is calculated from the feature maps extracted from the image by a pre-trained neural network.

Gramian matrices have been widely used in style transfer tasks, where the style of one image is transferred onto the content of another [3]. However, their potential for style similarity comparison in the context of plagiarism detection is also promising. By comparing the Gramian matrix of the flagged image with those of existing artworks in the database, we can assess the degree of stylistic similarity. If a high match is found even at this stage, the artwork is flagged for human inspection.

This two-pronged approach provides a comprehensive check for plagiarism, considering both the semantic content and stylistic elements of the artwork. The embedding-based similarity check efficiently narrows down potential matches, while the Gramian matrix analysis delves deeper into the stylistic nuances to identify potential cases of plagiarism that may not be apparent from content alone.

By implementing this rigorous plagiarism detection system, EVA Gallery aims to create a fair and transparent platform where artists can showcase their work with confidence, knowing that their creations are protected from unauthorized copying and misuse.

2 AI: Plagiarism Protection

Employ advanced techniques to detect and prevent plagiarism, ensuring the originality of uploaded artwork.

- **Embedding Space Lookup:** Artwork will be converted into numerical representations (embeddings) and compared within a vast database of existing art to identify potential copies or derivatives.
- **Gramian Matrix Similarity (Pending Research):**[1] A potential method that could further enhance plagiarism detection by analyzing the structural similarities between artworks.
- **Two-pronged Protection:** This system will guarantee that no modified artwork can be uploaded without proper attribution and permission, safeguarding artists' rights and fostering a fair creative environment.

[1] Further research is needed to validate the effectiveness of this method for plagiarism detection.

3 AI: Art Recommendation Engine

Provide personalized recommendations to users based on their preferences and interactions with the artwork.

- **Multi-Faceted Similarity:** Recommendations will consider both stylistic and content-based similarity to provide a diverse and engaging selection of artwork based on currently observed or displayed artwork.
- **Gramian-Based Style Recommendations:** Style-similar artwork will be evaluated using Gramian matrix similarity if proven viable.
- **Embedding-Based Content Recommendations:** Context-similar artwork will be displayed with respect to the typical similarity between embedded representations with MetaCLIP model.
- **Art Search:** Users will be able to input text that will be embedded and used to obtain images that most closely match the query.

3.1 Embedding-based content recommendations

EVA Gallery will take advantage of the power of image embeddings to provide content-based recommendations. By comparing the embedding of a user's observed artwork with the embeddings of other artworks in the database, we can identify pieces with similar semantic content, such as shared themes, subjects, or concepts. This enables us to recommend artworks that are similar to the currently observed artwork.

3.2 Mixed recommendations (embeddings and Gramian matrices)

To offer a diverse range of recommendations, we also consider a mixed approach that combines both embedding-based content similarity and Gramian matrix-based style similarity. This allows us to recommend artworks that share similar themes and visual styles with the observed artwork, providing a more comprehensive and engaging recommendation experience.

By incorporating both content and style information, we can cater to users who are interested in exploring artworks with similar subjects but different styles, or vice versa. This approach enhances the discovery aspect of the platform, exposing users to a wider array of artistic expressions.

3.3 Gramian-based style recommendations

For users who are primarily interested in visual styles, we offer recommendations based solely on Gramian matrix similarity. This allows us to identify artworks that share similar textures, colors, and patterns with the user's liked pieces, regardless of their semantic content.

This feature is particularly useful for users who are seeking inspiration for their own artistic creations or who are interested in exploring different artistic styles within a specific genre. By focusing on visual similarity, we can provide recommendations that corresponds best with the currently displayed artwork.

4 Data pipeline: Image upload and processing workflow

Upon uploading an image to EVA Gallery, the system generates its embedding using the MetaCLIP model. This embedding captures the semantic content of the image and serves as the primary basis for content-based similarity search and recommendations.

The generated embedding is compared against all existing embeddings in the vector database using efficient similarity search techniques like HNSW indexing. If a close match is found, indicating potential plagiarism, the system extracts a Gramian matrix from either the encoded representation feature maps or from pre-computed matrices stored in the database. This Gramian matrix is then compared against the Gramian matrix of the uploaded image to assess the degree of stylistic similarity. If a high match is detected, the image is flagged for human review to determine whether it constitutes plagiarism. This and the whole pipeline is visualized in Figure 1.2.

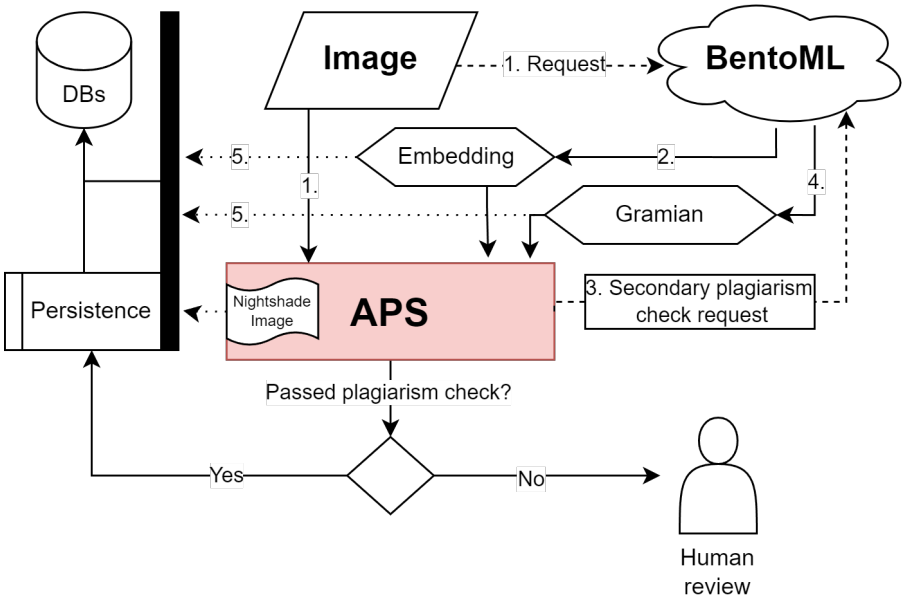

Figure 1.2: Integration schema of the data pipeline, in other words, the image flow through the connected components of the system and the related infrastructure.

Images that pass the plagiarism check are stored in the database along with their corresponding embeddings and Gramian matrices. Additionally, relevant metadata, including artist attribution, ownership details, and licensing terms, is associated with the image.

The artist has the option to mint the uploaded artwork as an NFT. The image is also processed with Nightshade, introducing subtle perturbations to render it useless for AI training while preserving visual fidelity. The Nightshade-processed image is stored as a separate copy in the database, linked to the original image via its unique UUID. This modified image is then displayed in the gallery, ensuring that visitors interact with the protected version while the original remains secure.

Chapter 2

System Requirements

1 AI

1. **Database Access:**
   - Access to a database of existing images (UUID identified).
   - Storage for modified images (Nightshade) alongside originals.
2. **Vector Database:**
   - Storage for embeddings and Gramian matrices.
   - HNSW indexing for efficient similarity search.
3. **Cloud Backend:**
   - Hosts MetaCLIP (embeddings) and ResNet (Gramian matrices) models.
   - Supports batched input.
   - Simple API (e.g., BentoML).
4. **Hardware Requirements:**
   - MVP: GPU with at least 24GB VRAM.
   - Deployment: Scalable GPU compute with load balancing and in-flight batching

Chapter 3

Technology stack

1 AI

1. **Database Access:**
   - **PostgreSQL:** A robust open-source relational database management system, offering support for structured data, ACID compliance, and scalability. Ideal for storing image metadata and maintaining associations between original and modified images.
2. **Vector Database:**
   - **pgvector+pgvectorscale+cube:** A PostgreSQL extension adding support for vector data types with similarity search using indexing algorithms, and matrix storage and search with cube. Cost-effective and scalable for storing and searching embeddings and Gramian matrices.
3. **Cloud Backend:**
   - **BentoML:** A flexible framework for building and deploying machine learning services. Simplifies packaging and deployment of models, allowing easy integration with various cloud providers, and supports batched input processing and API generation.
   - **Self-hosted:** Kubernetes provides control and flexibility in configuring the software environment and automating deployment and building. We can utilize our existing servers with a proper MLDevOps architecture to ensure CI and CD alongside AI models.
4. **Scalable GPU Compute:**
   - **Nvidia Triton Inference Server:** A high-performance inference server optimized for serving deep learning models. Supports dynamic batching and model ensemble, maximizing GPU utilization and throughput. Integrates with Kubernetes for seamless scaling and management.

Chapter 4

Current state of progress

This chapter addresses the project's current state and plans for the initial launch of EVA Gallery in the alpha/beta state. It will outline the features that are currently planned to be included and which are to be excluded initially, and the reasons for those decisions.

1 Artificial Intelligence

As we prepare for the initial launch of EVA Gallery in its alpha/beta state, the following key features have been prioritized for inclusion:

Ability to Find Similar Images Based on Currently Viewed Image: Users will be able to easily discover visually similar artwork by utilizing a feature that identifies and suggests images related to the one currently being viewed. This capability leverages advanced image recognition and comparison algorithms.

AI-Powered Image Search: The search functionality within EVA Gallery will be enhanced by embedding both text queries and images into a shared vector space. This allows users to perform more intuitive and accurate searches, finding artworks that match their textual descriptions or visual preferences.

Embedding-Based Plagiarism Detection: To protect artists' intellectual property, the platform will include an embedding-based plagiarism detection system. This system converts artwork into numerical embeddings and compares them against a database to identify potential copies or derivatives, ensuring that the originality of artwork is preserved.

NoAI and NoTrain Meta Tags: EVA Gallery will implement NoAI and NoTrain meta tags within its site metadata. These tags will serve as signals to scraping bots, indicating that the site's content is off-limits for AI training purposes. This proactive measure will help protect the artwork from being exploited by AI models without consent. item[Other Scraping Protections:] In addition to meta tags, EVA Gallery will employ various other scraping protection techniques to prevent unauthorized data extraction. These protections will be part of a broader strategy to safeguard artists' work and maintain the integrity of the platform.

The following features will instead be excluded due to time constraints or limitations outside of our control, for the alpha/beta launch:

Nightshade and Glaze AI-Training Protection: While these tools offer promising methods for preventing AI models from using protected artwork, their implementation requires access to specific model weights that are not currently publicly available. We have contacted the authors of these tools and are awaiting their response. Until these resources are accessible, these protections cannot be integrated into the platform.

Gramian-Based Similarity and Search: This advanced feature, which involves using Gramian matrices to evaluate and search for stylistic similarities between artworks, remains theoretical and unproven. Further research and validation are needed before it can be included in EVA Gallery's feature set. As such, it will not be available in the initial launch but may be considered for future updates.

Bibliography

1. SHAN, Shawn; DING, Wenxin; PASSANANTI, Josephine; WU, Stanley; ZHENG, Haitao; ZHAO, Ben Y. Nightshade: Prompt-Specific Poisoning Attacks on Text-to-Image Generative Models. In: *2024 IEEE Symposium on Security and Privacy (SP)*. IEEE Computer Society, 2024, pp. 212–212.
2. SHAN, Shawn; CRYAN, Jenna; WENGER, Emily; ZHENG, Haitao; HANOCKA, Rana; ZHAO, Ben Y. Glaze: Protecting Artists from Style Mimicry by {Text-to-Image} Models. In: *32nd USENIX Security Symposium (USENIX Security 23)*. 2023, pp. 2187–2204.
3. NICOLAS, Chung; XIE, Rong; SONG, Li; ZHANG, Wenjun. Improving Semantic Style Transfer Using Guided Gram Matrices. In: *Digital TV and Multimedia Communication: 15th International Forum, IFTC 2018, Shanghai, China, September 20–21, 2018, Revised Selected Papers 15*. Springer, 2019, pp. 169–183.