2026 IEEE 23rd Consumer Communications & Networking Conference (CCNC) | 979-8-3315-9673-6/26/$31.00 ©2026 IEEE | DOI: 10.1109/CCNC65079.2026.11366470

Provenance of AI-Generated Images: A Vector Similarity and Blockchain-based Approach

Jitendra Sharma*, Arthur Carvalho†, Suman Bhunia*
* Department of Computer Science and Software Engineering, Miami University, Oxford, Ohio
† Farmer School of Business, Miami University, Oxford, Ohio
Email: sharmaj2@miamioh.edu, arthur.carvalho@miamioh.edu, bhunias@miamioh.edu

***Abstract*—Rapid advancement in generative AI and large language models (LLMs) has enabled the creation of highly realistic and context-aware digital content. Multimodal systems such as ChatGPT with DALL-E integration and diffusion models like Stable Diffusion can now generate images that are often indistinguishable from human-created media, posing significant challenges for content authenticity and provenance. Ensuring the integrity and origin of digital data is critical for preserving trust, accountability, and legal compliance in digital ecosystems. This paper presents an embedding-based framework for detecting AI-generated images using vector similarity analysis. The proposed approach is based on the hypothesis that AI-generated images exhibit closer embedding proximity to other synthetic content, whereas human-created images cluster within distinct semantic regions. We validate this hypothesis by developing a prototype system that processes a diverse dataset of AI- and human-generated images using five benchmark embedding models. Furthermore we investigated whether perturbation with image, such as blurring or putting patches, would allow AI-generated images to confuse the system. Experimental results confirm the robustness of the method, showing that image perturbations have minimal impact on the embedding space, with altered images retaining high similarity to their originals.**

***Index Terms*—AI-generated image, generative AI, image detection, embedding, vector similarity, blockchain, provenance**

I. Introduction

Generative AI and large language models (LLMs) like ChatGPT have advanced rapidly, producing human-like text and realistic images from simple prompts. Diffusion models such as DALL-E, Imagen, and Stable Diffusion [1] now generate diverse visuals in seconds, making AI outputs increasingly indistinguishable from human-created content and challenging to detect.

The proliferation of AI-generated content raises concerns in creative and informational domains, as these models can imitate human artistry [2], [3] and devalue original work [4]. Notable incidents have fueled debates over AI's impact on art [5] and its misuse in generating fake images [6]. Existing detection methods often lack generalization across models [7], [8], while watermarking remains vulnerable [9], highlighting the need for more robust, adaptable solutions.

Recent studies have enhanced AI-image detection through forensic, feature-based, and hybrid approaches. Corvi et al. [10] analyzed diffusion model artifacts, noting detector performance dropped after compression or resizing. Ojha et al. [11] used CLIP-ViT embeddings for classification, improving generalization but struggling with heavy manipulations. Baraheem et al. [12] achieved 100% accuracy using EfficientNetB4 and CAMs, though requiring curated training data. Martin et al. [13] combined PRNU and error-level analysis with CNNs, gaining precision but with limited image format and adversarial robustness. These studies highlight the need for more generalizable and manipulation-resistant techniques.

Motivated by these limitations, we propose EmbedAIDetect—a training-free method that uses image embeddings and vector similarity to detect AI-generated content. The system employs a vision transformer-based embedding model with a vector database for comparison and integrates blockchain for provenance verification.

When an image is uploaded, its embedding is generated using the DINOv2 model and then compared against AI-generated images and human images reference sets. The closest match from each set is identified using cosine similarity. The closer match determines whether it is AI-generated or human-created. To verify the integrity of the result, the system computes a hash of the closest matched embedding and checks it against the corresponding smart contract. If a match is found on the blockchain, the classification is marked as verifiable; if not, it remains inference-based and unverifiable.

The main contributions of this research are:

- A resource-efficient, training-free system for AI image detection and provenance.
- A framework integrating vector databases, image embeddings, and blockchain for tamper-proof verification, with open-source code available [14].

The rest of the paper is organized as follows: Section II describes the system design; Section III evaluates the proposed system; and Section IV concludes the paper.

II. System Design

This section presents the architecture of the EmbedAIDetect system, which combines vector similarity search and blockchain verification to determine whether an image is AI-generated or human-created.

The system maintains two separate vector embedding databases: one for AI-generated images and the other for human-created images. When a new image is analyzed, its embedding is computed and compared against both databases to identify the closest match. Based on the cosine distance, it is determined whether the image is more likely to be generated by AI or created by a human. This approach is based on the

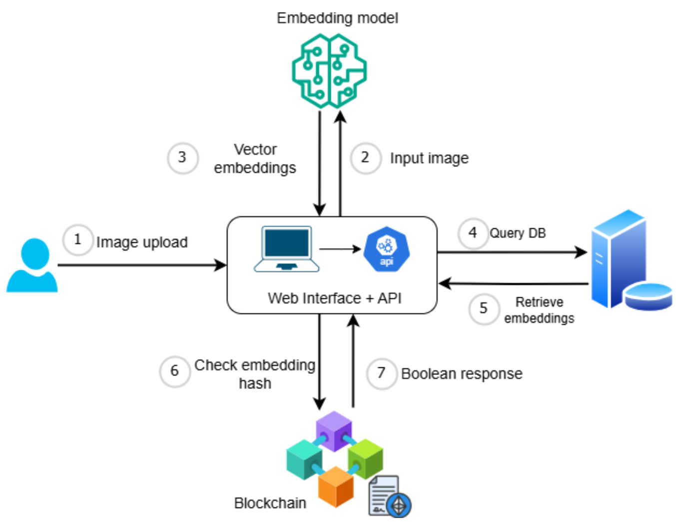


Fig. 1: System architecture for EmbedAIDetect

hypothesis that vector similarity in the embedding space can be used to distinguish between images produced by AI and those created by humans.

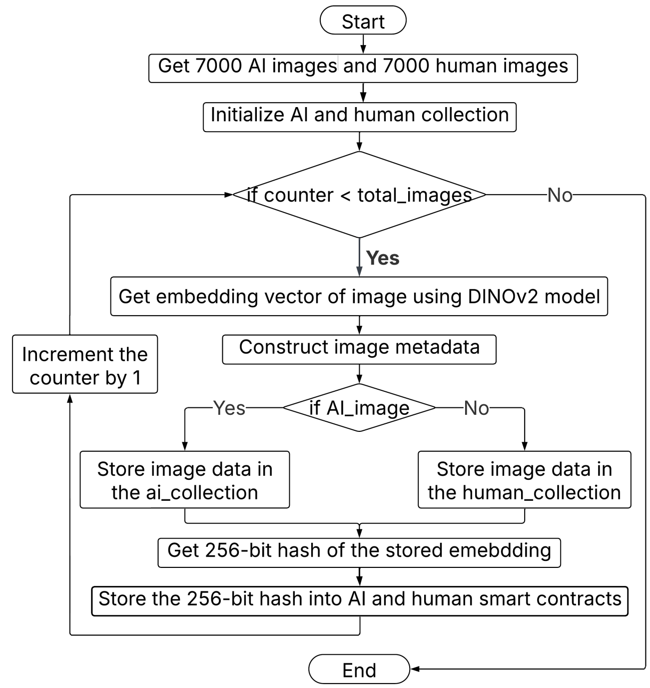
The architecture of the system is depicted in Figure 1, which shows the end-to-end pipeline in a labeled flow. The process begins when the user uploads an image through the web interface (Step 1). The image is then passed to the embedding model (Step 2), which generates a vector representation of the input (Step 3). This vector is used to query the vector database (Step 4), and the closest matching embeddings are retrieved (Step 5) to calculate cosine similarity. The hash of the image embedding is computed and checked against a blockchain smart contract to verify whether it has been previously registered (Step 6). The system then returns a final classification along with its verification status (Step 7).

A. *Provenance of Image Generated by AI or Human*

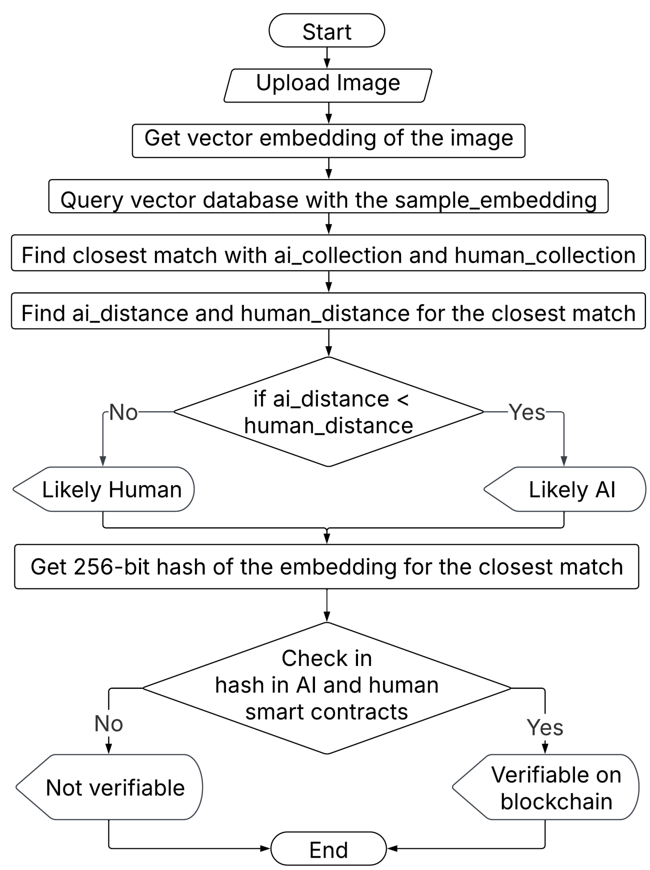
The proposed EmbedAIDetect system is designed as a unified framework that combines semantic similarity analysis and on-chain verification to determine the provenance of images, as illustrated in Figure 2. It integrates vector-based embedding classification with blockchain-backed validation to ensure both accurate inference and tamper-resistant verification.

The overall architecture is divided into two stages: application setup and application flow. The setup phase (Figure 2a) initializes two distinct ChromaDB collections and two blockchain smart contracts—`HashStorageAI` and `HashStorageHuman`—corresponding to AI-generated and human-created images. A dataset of 7,000 images per class is processed using the DINOv2 embedding model to generate high-dimensional feature vectors. These embeddings are stored in their respective ChromaDB collections for similarity search, while a 256-bit hash of each embedding is computed and recorded on the blockchain through the associated smart contracts.

The application logic is shown in the flow diagram (Figure 2b). When a new image is uploaded, the system computes its embedding and performs a similarity search across both

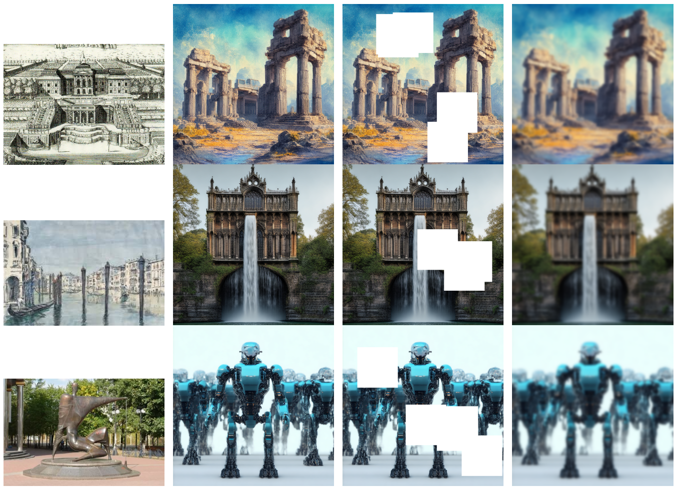


(a) Embeddings and 256-bit hash storage setup



(b) Vector similarity search and blockchain verification

Fig. 2: Application setup and data flowchart for the proposed framework of EmbedAIDetect



(a) (b) (c) (d)

Fig. 3: Sample images from human-art dataset and AI-generated image dataset with various modifications: (a) Human-generated art, (b) AI-generated image using prompt, (c) Images modified with multiple patches, (d) Blurred images

collections to identify the closest match. Based on the cosine distances, the system classifies the image as likely AI- or human-generated. It then verifies the result by hashing the closest matched embedding and checking for its presence on the blockchain. If the hash exists in the smart contract, the classification is marked as verifiable; otherwise, it remains inference-based and unverifiable.

III. Evaluation

EmbedAIDetect performance was evaluated through three studies: similarity-based classification, embedding robustness under adversarial conditions, and blockchain verification cost analysis. Experiments were performed on a system with an Intel Core i9-14900K CPU, 66 GB RAM, and an NVIDIA RTX 4090 GPU, while blockchain integration and gas usage were tested using Hardhat.

A. Vector similarity based classification

This experiment tests the effectiveness of vector embeddings and cosine similarity in distinguishing AI-generated from human-created images. A balanced dataset of 9,000 AI images (from Stable Diffusion 3.5 Medium at 512×512 resolution) and 9,000 human artworks (from Kaggle [15]) was used, excluding human faces. Sample images and manipulations are shown in Figure 3.

The methodology uses a classification pipeline based on vector embeddings from pre-trained models. Embeddings of AI-generated and human images are stored in a Pinecone vector database with an 80/20 training–testing split. For each test image, the nearest neighbor is retrieved using cosine similarity. Experiments were conducted using five models—OpenAI CLIP, Google ViT, Facebook DINOv2, Microsoft ResNet-50, and Apple AIMv2 with separate indexes maintained for each.

TABLE I: Performance Metrics Across Embedding Models

Models	**Embedding Dimension**	**Precision**	**Recall**	**Accuracy**
CLIP	512	0.9935	0.998	0.9951
Google VIT	768	0.988	0.987	0.9856
DinoV2	768	0.9881	0.9945	0.9899
ResNet 50	2048	0.9898	0.826	0.8948
AIMv2	1024	0.992	0.994	0.9919

TABLE II: Model Accuracy (%) for Patch Overlays and Resize

Models	**Embedding Dimension**	**1 Patch**	**3-5 Patches**	**Resize**
CLIP	512	97.26	73.34	98.04
Google VIT	768	99.88	91.62	98.98
DinoV2	768	100.00	99.62	99.94
ResNet-50	2048	99.89	84.02	98.47
AIMV2	1024	99.96	92.59	99.81

TABLE III: Model Accuracy (%) Under Blur Intensities

Models	**Embedding Dimension**	**Blur Intensity**			
		20%	**40%**	**60%**	**80%**
CLIP	512	96.68	80.96	48.12	30.02
Google ViT	768	99.86	96.00	84.17	67.06
DinoV2	768	100.00	99.87	96.86	88.44
ResNet-50	2048	98.52	74.18	44.36	24.67
AIMv2	1024	99.65	96.81	90.08	80.13

As shown in Table I, all models performed exceptionally well. CLIP achieved the highest accuracy (99.51%) with near-perfect precision and recall, while AIMv2 showed consistent results. ResNet-50 exhibited slightly lower recall but maintained high precision. Except for ResNet-50, all models exceeded 98% accuracy with minimal misclassification.

B. Benchmarking different embedding models

This experiment evaluates the robustness of embedding models against image manipulations. Modified AI-generated images were compared with their original versions to assess how well embeddings preserve similarity under distortions. Two types of alterations were tested: geometric (white patch overlays, resolution reduction) and blur transformations.

Using the same 9,000 AI-generated images, six modified sets were created, and embeddings were generated with the five models from Experiment 1. Each modified image was compared with its original using cosine similarity in Pinecone, and the accuracy was calculated as $\frac{\text{Number of Correct Matches}}{\text{Total Number of Images}}$ where a correct match indicates that the embedding of the modified image was closest to its original.

As shown in Table II, DINOv2 showed the highest robustness to geometric variations, achieving 100% accuracy with single patches and 99.94% at reduced resolution (128×128). Multiple patch overlays were more challenging, reducing CLIP's accuracy to 73.34%.

Blur analysis (Table III) revealed an inverse relationship between blur intensity and accuracy. At mild blur (20%), most models exceeded 99%, with DINOv2 achieving perfect accuracy. Even at 80% blur, DINOv2 maintained 88.44%, far surpassing other models, while ResNet-50 dropped to 24.67%.

The Results indicate that AI-generated images form consistent, identifiable patterns in embedding space. Transformer-based models outperform CNNs like ResNet-50, and ro-

TABLE IV: Gas Usage Statistics for Hash Storage Using `uint256` and `string`

Statistic	Hash (uint256)	Hash (string)
Count	9,000 transactions	9,000 transactions
Mean	36,207 gas	97,667 gas
Median	36,184 gas	97,667 gas
Minimum	21,528 gas	97,667 gas
Maximum	51,228 gas	97,667 gas

bustness depends more on architecture than on embedding dimensionality.

C. Smart Contract Gas Usage Estimation

This experiment compared gas consumption when storing embedding hashes on Ethereum using two Solidity data types: `uint256` and `string`. Two contracts, `HashStorageInt` and `HashStorageStr`, were implemented and tested in a Hardhat environment, each containing functions to store and verify hashes. The summarized statistics are provided in Table IV. Gas usage for the `storeHash` operation was recorded, showing that `uint256` storage is about 2.7 times more efficient than `string`.

D. End-to-End Prototype Evaluation

To assess the practical performance of our AI image provenance and detection system, we developed and deployed a working prototype capable of classifying uploaded images as AI-generated or human. Tests covered diverse inputs, including StyleGAN-generated faces and non-facial images, to assess accuracy and identify potential limitations.

1) Blockchain Verification Evaluation: Only a few hundred (approximately 500) image embedding hashes could be stored on the Ethereum Sepolia testnet due to network constraints, including gas fees, transaction limits, nonce conflicts, and rate limiting by the node provider (*Alchemy*). To evaluate functionality at scale, the Hardhat local blockchain was used to simulate storage and verification. While the setup confirmed the system's ability to match and verify embeddings on the blockchain, this evaluation also revealed limitations in scalability and the cost of using blockchain with large datasets.

2) Similarity-based Classification Evaluation: This evaluation examined embedding-based classification using vector similarity search. To improve facial image detection, 10,000 StyleGAN AI-generated faces from `thispersondoesnotexist` were added, raising synthetic face accuracy above 90% on 2,500 tests but introducing bias, as real faces were often misclassified. Adding 1,441 real faces from the Chicago Face Database (CFD) [16] improved balance but distinguishing realistic faces remained challenging.

Tests on external datasets further validated performance. On CIFAKE [17], the system correctly classified 61.9% of real and 47.9% of fake images, showing moderate effectiveness for diverse content. On the 140k Real and Fake Faces dataset [18], 94.5% of real faces were identified correctly, but only 2.4% of fake ones. These results indicate that while embedding-based similarity performs well on general imagery, it remains limited against highly realistic facial forgeries.

IV. Conclusion

This paper presents a novel concept for detecting and verifying the provenance of AI-generated images using embedding-based similarity and blockchain verification. The findings show that, despite extensive manipulations, image semantics remains essentially unchanged, allowing for identification through vector similarity. DINOv2 demonstrated strong robustness under such conditions, making it suitable for modified-image scenarios. However, the effectiveness of the approach is influenced by the diversity and quality of the dataset. Results indicate that system accuracy can be improved with a more representative dataset, though trade-offs between model generalization and performance must be considered when selecting pre-trained embedding models. Future work will explore decentralized storage (e.g., IPFS) and broader integration to standardize AI image provenance and enhance trust in digital media.

References

[1] S. Bengesi *et al.*, "Advancements in Generative AI: A Comprehensive Review of GANs, GPT, Autoencoders, Diffusion Model, and Transformers," *IEEE Access*, 2024.

[2] Z. Winn, "Pushing the Frontiers of Art and Technology with Generative AI." https://news.mit.edu/2023/pushing-frontiers-art-technology-generative-ai-1129, 2023.

[3] S. M. Kalita and A. Iyer, "How Generative AI Could Disrupt Creative Work." https://hbr.org/2023/04/how-generative-ai-could-disrupt-creative-work, 2023.

[4] A. Madhu, "Survey Reveals 9 out of 10 Artists Believe Current Copyright Laws are Outdated in the Age of Generative AI Technology." https://bookanartist.co/blog/2023-artists-survey-on-ai-technology/, 2023.

[5] K. Roose, "An A.I.-Generated Picture Won an Art Prize. Artists Aren't Happy," Sept. 2022.

[6] A. Lajka and P. Marcelo, "Fake AI images of Putin, Trump being arrested spread online," 2023.

[7] M. Zhang, "AI or Not is a Free Web App that Claims to Detect AI-Generated Photos." https://petapixel.com/2023/06/16/ai-or-not-is-a-free-web-app-that-claims-to-detect-ai-generated-photos/, 2023.

[8] N. Sandotra and B. Arora, "A comprehensive evaluation of feature-based AI techniques for deepfake detection," *Neural Computing and Applications*, vol. 36, no. 8, pp. 3859–3887, 2024.

[9] Z. Jiang, J. Zhang, and N. Z. Gong, "Evading Watermark based Detection of AI-Generated Content," in *ACM SIGSAC Conference on Computer and Communications Security*, pp. 1168–1181, 2023.

[10] R. Corvi *et al.*, "On the detection of synthetic images generated by diffusion models," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5, 2023.

[11] U. Ojha, Y. Li, and Y. J. Lee, "Towards Universal Fake Image Detectors that Generalize Across Generative Models," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 24480–24489, 2023.

[12] S. S. Baraheem and T. V. Nguyen, "AI vs. AI: Can AI Detect AI-Generated Images?," *Journal of Imaging*, vol. 9, no. 10, p. 199, 2023.

[13] F. Martin-Rodriguez, R. Garcia-Mojon, and M. Fernandez-Barciela, "Detection of AI-Created Images Using Pixel-Wise Feature Extraction and Convolutional Neural Networks," *Sensors*, vol. 23, no. 22, 2023.

[14] "AiOrNot Github Repository." https://github.com/sbhunia/AiOrNot.

[15] TheDownhill, "Art Images – Drawings, Paintings, Sculptures, Engravings." https://www.kaggle.com/datasets/thedownhill/art-images-drawings-painting-sculpture-engraving, 2018.

[16] D. S. Ma, J. Correll, and B. Wittenbrink, "The Chicago Face Database: A Free Stimulus Set of Faces and Norming Data," *Behavior Research Methods*, vol. 47, pp. 1122–1135, 2015.

[17] F. Birdy, "CIFAKE: Real and AI-Generated Synthetic Images." https://www.kaggle.com/datasets/birdy654/cifake-real-and-ai-generated-synthetic-images, 2023.

[18] X. Huang, "140K Real and Fake Faces." https://www.kaggle.com/datasets/xhlulu/140k-real-and-fake-faces, 2021.