Revolutionizing Fashion Design with Multimodal Generative AI And Blockchain-Enabled Ownership

Sanika Kulkarni[1] Balaji Chaugule[2]
[1]*Student, Dept. of Data Science, Zeal College of Engineering and Research*
[2]*Professor, Zeal College of Engineering and Research*

***Abstract*—The fashion industry is increasingly shaped by consumer demand for personalized, sustainable, and distinctive designs; however, traditional design processes remain time-consuming and are often unable to meet this demand efficiently. Existing artificial intelligence approaches—particularly Generative Adversarial Networks (GANs)—exhibit limitations such as mode collapse, which results in repetitive and non-diverse outputs. To address these challenges, this paper introduces Style Sensei, a multimodal generative AI–based fashion assistant integrated with blockchain technology. Style Sensei supports text-to-image generation, image-to-image transformation, and sketch-to-image conversion, enabling designers and consumers to explore virtually limitless creative possibilities. Unlike conventional GAN-based systems, Style Sensei leverages diffusion-based generative models to ensure diversity and novelty in fashion outputs. In addition, a conversational AI chatbot delivers personalized outfit recommendations, trend insights, and styling guidance, thereby enhancing user engagement. To guarantee authenticity and ownership, Style Sensei integrates blockchain and Inter Planetary File System (IPFS) technologies, ensuring that each generated design is securely stored and verifiably linked to its creator. Smart contracts record design ownership by associating blockchain wallet addresses with IPFS content identifiers, preventing misuse and plagiarism of digital fashion assets. Experimental evaluation demonstrates the system's effectiveness in producing high- quality, unique designs while ensuring trust, transparency, and creative freedom. Overall, this work highlights the potential of combining multimodal generative AI with blockchain to revolutionize the fashion industry by creating a secure, diverse, and personalized design ecosystem.**

***Index Terms*—Generative Artificial Intelligence, Multimodal Learning, Fashion Design, Stable Diffusion, Blockchain, IPFS, Smart Contracts**

I. INTRODUCTION

Consumer demand for personalized, sustainable, and distinctive designs is transforming the fashion industry. Traditional design workflows that rely heavily on manual sketching, prototyping, and iterative refinement are resource-intensive and often unable to adapt quickly to changing market dynamics. Consequently, designers and brands are increasingly turning to artificial intelligence (AI) to augment creativity, accelerate production cycles, and meet evolving customer expectations. However, current AI-based solutions face significant limitations that hinder their effectiveness in real-world fashion applications.[1] Generative Adversarial Networks (GANs), one of the most widely adopted approaches for image synthesis, have demonstrated notable success in generating fashion images; nevertheless, they continue to suffer from challenges such as mode collapse, instability during training, and limited control over creative outputs.

This limitation results in repetitive patterns and reduced diversity, thereby constraining the creative exploration that designers require. Furthermore, many existing AI tools are unimodal, relying solely on text or image inputs and consequently excluding sketches or hybrid multimodal inputs that are essential to the design process. These shortcomings underscore the need for more robust, versatile, and scalable AI-driven fashion-design frameworks.

Recent advancements in diffusion-based generative models have opened new opportunities to overcome the limitations of Generative Adversarial Networks (GANs) [2], particularly with respect to diversity, stability, and realism. These models support powerful multimodal capabilities— including text-to-image, image-to-image, and sketch-to- image generation—

enabling designers to transform simple prompts, existing outfits, or rough sketches into professional-quality fashion illustrations. When integrated with conversational AI, such systems extend beyond design assistance to deliver personalized outfit recommendations, trend insights, and interactive style guidance, making them especially well-suited to creative industries such as fashion.[3]

However, the growing prevalence of AI-generated content introduces substantial challenges concerning authenticity, plagiarism, and ownership of digital assets. To address these issues, blockchain technology provides a decentralized and tamper-proof mechanism for securing intellectual property rights. By integrating Inter Planetary File System (IPFS) storage with blockchain smart contracts, fashion designs can be permanently linked to their rightful creators, thereby ensuring verifiable ownership and preventing misuse. Within this context, this paper introduces Style Sensei, a generative AI–powered multimodal fashion assistant that unites diffusion-based creativity with blockchain-enabled provenance. By combining diverse generative capabilities with secure ownership verification, Style Sensei seeks to redefine trust, personalization, and creativity in the fashion ecosystem.

II. RELATED WORK

Recent advancements in generative AI—particularly latent diffusion models—have transformed fashion design by enabling the creation of realistic and diverse outfits. Frameworks such as Multimodal Garment Designer [4] and FashionSD-X demonstrate how text, sketches, and body poses can be combined to generate coherent, human-centric fashion imagery. These studies highlight the superiority of diffusion-based methods over GANs, as they circumvent issues such as mode collapse and support flexible, multimodal creativity within design workflows.

Alongside advances in generative capabilities, research has increasingly emphasized enhancing human–AI collaboration to improve creative efficiency. For example, the HAIGEN system [5] combines cloud-based generation with locally deployed sketch libraries and style-aware tools, enabling designers to maintain privacy while benefiting from AI-assisted ideation. Simultaneously, blockchain and NFT technologies [6] have been explored to secure ownership of AI-generated designs, introducing “phygital” fashion that merges digital and physical goods.

Decentralized storage solutions such as IPFS further reinforce provenance and ensure long-term accessibility of creative assets. More recent studies have extended generative workflows to support greater control and multi- input pipelines. Models such as Text Control [7] and Sketch2Cloth enable sketch-guided garment generation, while projects like from sketch to reality and frameworks incorporating DALL·E 2, ControlNet, T2IAdapter, and CLIP [8] demonstrate personalized and multimodal consistency in fashion synthesis. Despite these advances, existing systems remain fragmented—tending to focus either on creative generation or on ownership verification. This gap underscores the need for an integrated solution such as Style Sensei, which unifies multimodal generative AI with blockchain-based ownership to ensure both limitless creativity and secure provenance of digital fashion assets.

III. METHODOLOGY

1. Text-to-Image Generation Module

The text-to-image generation module of the proposed system is built on Stable Diffusion, an open-source, diffusion-based generative model implemented via the Hugging Face Diffusers library. The workflow begins when a user submits a natural-language prompt through the web interface, which is transmitted to the Flask backend via an HTTP request. The model employs a CLIP text encoder [9] to transform the textual input into embeddings that guide the generative process. Starting from a field of random noise in latent space, a U-Net–based denoising network iteratively refines the representation under the influence of the text embeddings, progressively shaping semantically coherent visual content. Once the diffusion process is complete, a Variational Autoencoder (VAE) decoder reconstructs the final high-resolution image from the latent representation. This pipeline enables the system to synthesize realistic and contextually relevant images aligned with the input description, forming the foundation for subsequent storage, ownership validation, and export processes in the overall platform.[10]

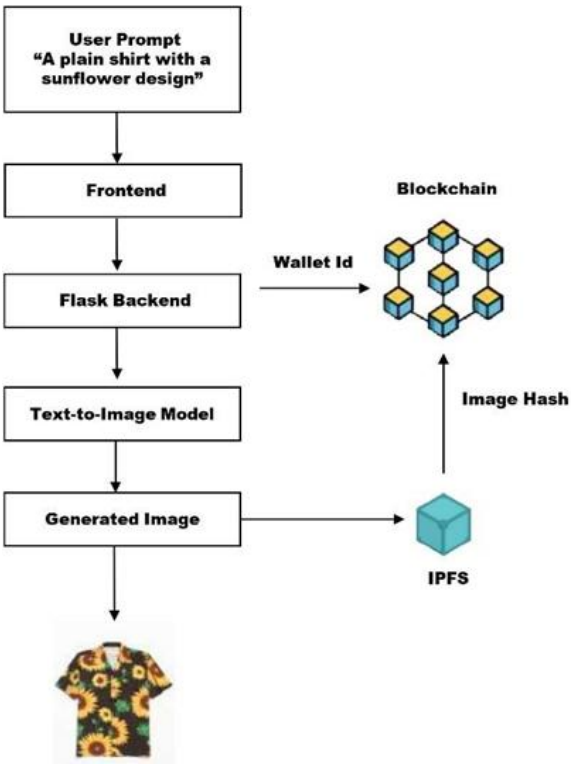

Figure 1: Text-to-Image Generation

2. Image-to-Image Generation Module

This module employs the Stable Diffusion Img2Img pipeline from Hugging Face's Diffusers library, enabling the transformation and refinement of input images guided by textual prompts. The process begins when a user uploads an initial reference image through the frontend, which is then transmitted to the Flask backend along with a descriptive prompt.[10] The pipeline introduces a controlled amount of noise to the input image and subsequently employs a U-Net diffusion model—conditioned on the new text embeddings provided by the CLIP encoder—to iteratively denoise and reconstruct the image. A DDIM scheduler provides deterministic and fine-grained control over the denoising trajectory, ensuring both fidelity to the original structure and adaptability to the new description. In the final stage, the Variational Autoencoder (VAE) decoder reconstructs the refined image from latent space into pixel space. This approach preserves the structural essence of the original image while integrating new stylistic or semantic features, making it particularly suitable for fashion design modifications and creative reinterpretations of existing concepts.

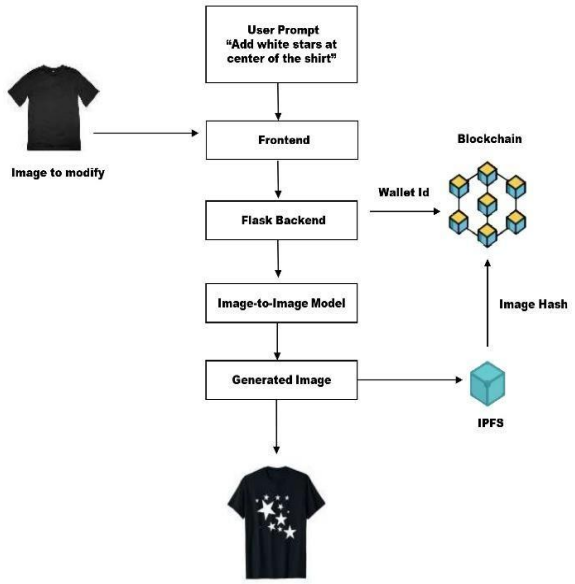

Figure 2: Image-to-Image Generation

3. Sketch-to-Image Generation Module

The sketch-to-image generation module extends the Stable Diffusion framework with ControlNet, enabling the system to synthesize realistic fashion designs from hand-drawn sketches. The workflow begins when a user uploads a sketch, which is preprocessed using OpenCV to extract clean binary edges through adaptive thresholding, ensuring that only the essential contours of the drawing are preserved. This processed sketch, together with the user's descriptive text prompt, is passed into the Stable Diffusion Control Net Pipeline integrated with the lllyasviel/sd-controlnet-scribble model. ControlNet functions as a structural constraint that anchors the diffusion process to the sketch's outlines, while the CLIP text encoder provides semantic guidance to shape textures, fabrics, and stylistic details.[11] The U-Net diffusion model then iteratively denoises the latent representations, combining both structural and textual cues, before the Variational Autoencoder (VAE) decoder reconstructs the final high-resolution image. This integration of sketch conditioning and prompt guidance enables the system to transform rough outlines into polished, detailed fashion designs, aligning user creativity with AI- driven refinement.

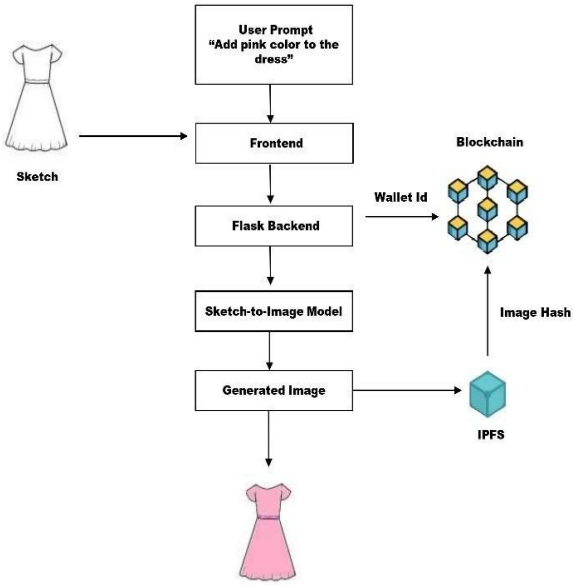

Figure 3: Sketch-to-Image Generation

The overall architecture of StyleSensei follows a client– server paradigm, combining a lightweight, web-based frontend with a Flask-powered backend that orchestrates AI pipelines, blockchain interactions, and file management. The frontend—built with HTML, Bootstrap, and JavaScript— serves as the user interface, allowing designers and consumers to input text descriptions, upload outfit images, or submit hand-drawn sketches. It also integrates MetaMask via Ethers.js to securely connect user Ethereum wallets for ownership verification. The frontend communicates with the backend through RESTful APIs, transmitting prompts, images, and wallet information to initiate the appropriate processing pipelines.

At the backend, Flask acts as the central controller, routing user requests to specialized AI modules and blockchain services. The AI design engine is powered by Stable Diffusion models from the Hugging Face Diffusers library, supporting three distinct generative modes: text-to- image, image-to-image, and sketch-to-image (with ControlNet for structural guidance). These modules leverage PyTorch for model inference with CUDA acceleration where available. Supporting libraries such as OpenCV and PIL handle preprocessing tasks such as edge detection, image normalization, and format conversion, ensuring clean and consistent inputs to the diffusion models. The generated images are stored locally before being uploaded to IPFS, providing decentralized and permanent storage of digital fashion assets.[12]

To ensure secure authorship and ownership, the system integrates a blockchain verification layer. Once a design is finalized, its corresponding IPFS Content Identifier (CID) is retrieved and bound to the user's Ethereum wallet address via a Solidity-based smart contract deployed on the Ethereum network.[12] This transaction immutably records ownership on-chain, while a ReportLab-generated PDF certificate consolidates the wallet address, CID, and design preview into a verifiable proof of ownership. Together, these components establish a unified ecosystem where design ideation, AI-driven generation, iterative refinement, decentralized storage, and blockchain-backed authorship verification are seamlessly interconnected.

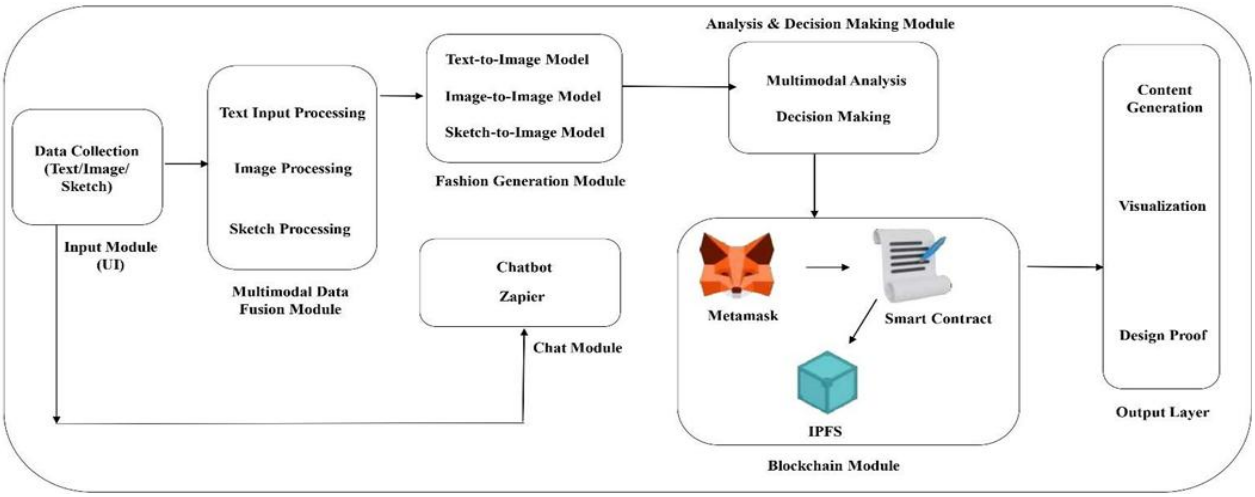

Figure 4: System Architecture

IV. EXPERIMENTS AND RESULTS

To evaluate the Style Sensei framework, experiments were conducted on its core generative pipelines—text-to-image, image-to-image, and sketch-to-image—by measuring execution time, CPU utilization, and RAM consumption. In addition, the impact of prompt complexity on the text-to-image pipeline was assessed by comparing short and long descriptive prompts, thereby highlighting how prompt length influences processing time and system resource usage. Collectively, these experiments provide insights into the computational performance and operational efficiency of the proposed system.

Table 1: System Performance

Pipeline	Execution Time (sec)	Observed Impact on Resources
Text-to-Image	157.7 (≈ 2m 37s)	Moderate CPU, High RAM usage
Image-to-Image	118.6 (≈ 1m 59s)	High CPU, Moderate RAM usage
Sketch-to-Image	146.7 (≈ 24m 23s)	Very high time, High RAM usage

Table 2: Impact of Prompt Complexity on Performance

Prompt Type	Execution Time (sec)	Resource Usage (CPU%, RAM%)
Short Prompt	152.37	CPU: 54.0% RAM: 95.5%
Long Prompt	161.36	CPU: 58.2% RAM: 92.4%

V. CONCLUSION

This research presents StyleSensei, a multimodal AI-powered fashion assistant designed to redefine the creative workflow for fashion designers and consumers. By integrating Stable Diffusion–based generative models with blockchain- backed ownership verification, StyleSensei bridges the gap among design ideation, visualization, and secure authorship. Unlike conventional approaches that rely on single-mode input or GAN-based models prone to repetitive outputs, StyleSensei enables a seamless design process across text-to-image, image- to-image, and sketch-to-image modalities, offering unparalleled flexibility and creative exploration.[14]

Beyond design generation, the system ensures trust and transparency by storing each creation on IPFS and recording ownership on the Ethereum blockchain via smart contracts. This dual-layer protection not only guarantees provenance but also empowers designers with verifiable proof of authorship through automatically generated certificates. Furthermore, the integration of a chatbot-driven recommendation system personalizes user experiences, making fashion ideation more interactive, efficient, and user-centered. In summary, StyleSensei demonstrates how the convergence of generative AI and blockchain technology can support the evolving demands of the fashion industry by offering faster prototyping, limitless ideation, and secure digital ownership. The system establishes a foundation for the future of digital fashion co-creation, where creativity is amplified by AI and protected by decentralized technologies.

VI. FUTURE SCOPE

The proposed StyleSensei framework demonstrates the potential of multimodal generative AI and blockchain in revolutionizing digital fashion design, with several opportunities for future enhancement. By integrating next- generation diffusion models (e.g., Stable Diffusion XL or fine-tuned fashion-specific models), the system could achieve greater realism, fabric detail, and cultural diversity. In addition, the inclusion of 3D garment generation and virtual try-on modules would bridge the gap between digital prototypes and real-world fashion by enabling accurate visualization on avatars. Personalization can be further advanced through reinforcement learning and user feedback loops, allowing the assistant to dynamically adapt to evolving preferences and fashion trends.[14]

On the blockchain side, expanding beyond Ethereum into multi-chain interoperability (e.g., Polygon, Solana) would reduce costs and improve accessibility, while integration with NFT marketplaces could empower designers to monetize their creations. Moreover, StyleSensei could evolve into a collaborative digital ecosystem where designers, consumers, and brands co-create, share, and manage digital fashion assets securely and transparently, fostering innovation and inclusivity across the fashion industry

REFERENCES

[1] Xiaoling Gu et al. ‘‘Fashion analysis and understanding with artificial intelligence’’. In: Information Processing & Management 57.5 (2020), p. 102276

[2] Ian Goodfellow et al. ‘‘Generative adversarial networks’’ In: Communications of the ACM 63.11 (2020), pp. 139–144

[3] Guo, Ziyue, Zongyang Zhu, "AI-Assisted Fashion Design: A Review," IEEE, 2023.

[4] A. Baldrati, D. Morelli, G. Cartella, M. Cornia, M. Bertini, and R. Cucchiara, “Multimodal Garment Designer: Human- Centric Latent Diffusion Models for Fashion Image Editing,” arXiv preprint arXiv:2304.02051, Apr. 2023, revised Aug. 2023.

[5] J. Jiang, D. Wu, H. Deng, Y. Long, W. Tang, X. Li, C. Liu, Z. Jin, W. Zhang, and T. Qi, “HAIGEN: Towards Human-AI Collaboration for Facilitating Creativity and Style Generation in Fashion Design,” arXiv preprint arXiv:2408.00855, August 2024. DOI: 10.48550/arXiv.2408.00855

[6] N. A. Volodeva, “Blockchain in fashion industry: Practice, prospects and challenges of NFT technology,” in Proceedings of the International Scientific Conference “Scientific Research of the SCO Countries: Synergy and Integration”, Beijing, China, May 2024, doi: 10.34660/INF.2024.63.41.166.

[7] Y. Zhang, T. Zhang, and H. Xie, “TexControl: Sketch- Based Two-Stage Fashion Image Generation Using Diffusion Model,” arXiv preprint arXiv:2405.04675, May 2024. DOI: 10.48550/arXiv.2405.04675

[8] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. Unet: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18thInternational Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, pages 234–241.Springer, 2015

[9] Xingran Zhou et al. ‘‘Text guided person image synthesis’’. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019, pp. 3663–3672

[10] Kulkarni, S., Balbudhe, K. S., Rasal, P., Mahankar, A., & Biradar, A. (2024, April). AI artisan in fashion design. TIJER – International Research Journal, 11(4), a903–a910. ISSN 2349-9249. Pune, Maharashtra, India

[11] Bansal, Manya, David Wang,” Sketch 2 Fashion: Generating clothing visualization from sketches”, “Stanford CS230," [Online].

[12] N. A. Volodeva, “Blockchain in fashion industry: Practice, prospects and challenges of NFT technology,” in Proceedings of the International Scientific Conference “Scientific Research of the SCO Countries: Synergy and Integration”, Beijing, China, May 2024, doi: 10.34660/INF.2024.63.41.166.

[13] Jascha Sohl-Dickstein, Eric A. Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using none quilibrium thermodynamics. CoRR, abs/1503.03585, 2015.1.

[14] B. Chaugule and S. Kulkarni, “Frontiers of Multimodal Generative AI: Efficiency, Adaptability, and Real-World Applications,” Int. J. Innov. Res. Technol., vol. 11, no. 8, pp. 770–778, Jan. 2025. ISSN: 2349-6002