Journal of Research in Computer Science and Engineering
Volume 11, Issue 1, January-April 2026
ISSN: 2457-0818 (Online)

AI VOGUE: A Multi-Modal Textile E-Platform (MMT-EP) Integrating Generative AI, AR Visualization, and P2P Commerce

Hemanth Harishetty[1], Rahul Kulkarni[2], Sakshi Shivapujimath[3], Soumya Halki[4], Priyanka Desurkar[5]

Student[1, 2, 3, 4], Assistant Professor[5]

Department of Computer Science and Engineering

Jain College of Engineering and Research, Belagavi

Corresponding Authors' Email: *hemanthharashetty7@gmail.com[1]*

DOI: *https://doi.org/10.5281/zenodo.19144061*

ABSTRACT

The current e-commerce landscape is highly fragmented, forcing users to switch between multiple platforms to meet different needs such as retail shopping, apparel rentals, and custom design services. This paper introduces AI VOGUE, a unified Multi-Modal Textile E-Platform (MMT-EP) designed to eliminate these inefficiencies through a cohesive Systemic Integration Architecture. Unlike conventional solutions that offer isolated features, AI VOGUE seamlessly integrates Generative AI–driven Virtual Try-On (VTO), AR-based accessory visualization, LLM-powered conversational styling, and a Multi-Source Aggregation Engine into one streamlined experience. The platform employs a hybrid technology stack, using Three.js for high-fidelity 2D and AR rendering, MediaPipe and Google ARCore for real-time accessory tracking, the Groq API for ultra-low-latency intelligent responses, and Appwrite as a scalable serverless backend. A major contribution of this work is the development of a hybrid VTO pipeline that combines the Gemini API with the Nano Banana Model to produce photorealistic garment transfer results, along with the creation of a decentralized "Creative Threads" marketplace that supports peer-to-peer designer interactions. *Overall, this research shows that integrating diverse computational models within a unified, user-centric system significantly enhances user engagement while democratizing access to advanced fashion technologies.*

KEYWORDS:- *Multi-Modal E-Platform, Generative AI, Virtual Try-On, Augmented Reality, Serverless Architecture, Conversa- tional Commerce.*

INTRODUCTION

The contemporary digital fashion ecosystem is increas- ingly burdened by structural fragmentation and isolated ser- vice models. Most existing platforms are highly domain-specific—focusing solely on mass retail, luxury rentals, or bespoke tailoring—which forces consumers to navigate a disconnected collection of services to meet their different sar- torial needs. This logistical inefficiency is further intensified by the persistent "experience gap" found in online shopping. Because platforms rely heavily on static, two-dimensional images, they fail to clearly convey important factors like fit, drape, and overall aesthetic compatibility. This limitation not only contributes to high return rates but also gradually weak- ens consumer confidence. As a result, the industry now faces a crucial need to move beyond these isolated approaches and shift toward a unified architecture that brings together diverse inventories while offering immersive, realistic visualization. The primary impetus for this research is to resolve this "fragmentation-experience" dichotomy. We argue that an ef- fective solution must go beyond simple aggregation; it calls for the creation of a holistic ecosystem that brings together diverse commerce models while also delivering advanced, AI-driven decision support systems. Such an environment needs to combine predictive styling intelligence with high-fidelity visual simulation in order to recreate the tactile, in teractive qualities of physical retail.

To address these systemic challenges, we propose the Multi-Modal Textile E-Platform (MMT-EP) Architecture, re- alized through the implementation of the AI VOGUE system. This framework marks a meaningful shift from isolated utili- ties to a fully cohesive service orchestration model. Our spe cific technical contributions to the field are threefold:

1. **Unified Service Orchestration:** We introduce a novel integration logic that seamlessly combines a Rental Hub ("Style Lease"), a Peer-to-Peer Custom Designer Market- place ("Creative Threads"), and an aggregated retail search engine into a single, state-managed workflow. This stream- lined orchestration removes the cognitive burden typically associated with switching between multiple platforms.
2. **Hybrid Intelligent Layer:** We propose a strategic hybridization of computational models, integrating Groq's Llama 3.3 for ultra-low-latency, context-aware styling rea- soning with

the Google Gemini API for high-fidelity gener- ative visual synthesis. This dual-engine setup optimizes the balance between fast conversational response and high visual realism.

3. **Scalable Serverless Infrastructure:** We demonstrate a robust deployment strategy that uses Appwrite as a Backend- as-a-Service (BaaS) backbone. This architecture efficiently handles complex, multi-modal user sessions and state tran- sitions while avoiding the operational overhead of tradi- tional monolithic servers, ensuring strong scalability and high availability.

LITERATURE REVIEW

Virtual and Augmented Try-on (VTO) Paradigms

The field of Virtual Try-On (VTO) has long been defined by a clear trade-off between computational efficiency and visual fidelity. Traditional VTO systems typically use geometric warping methods and basic 2D overlays to place garments onto user avatars. Although these approaches are computa- tionally lightweight, they often fall short in accurately rep- resenting complex fabric physics, drape behavior, and re- alistic lighting, which results in noticeably low visual real- ism. On the other hand, newer techniques that leverage Gen- erative Adversarial Networks (GANs) and diffusion models have greatly improved visual quality, enabling highly pho- torealistic outputs. Despite this progress, such deep learn- ing–based methods often struggle with high inference la- tency, making them unsuitable for real-time, web-based user interactions[4]. This work addresses this divide by introducing a Hybrid VTO Pipeline that intentionally pairs the rendering speed of Three.js with the generative in-painting capabilities of the Gemini API. This architecture effectively balances the need for real-time responsiveness to maintain user engage- ment with the high level of photorealistic detail required for accurate fit evaluation.

B. Contextual LLM Integration in E-commerce

Recommender systems in e-commerce have traditionally re- lied on collaborative filtering algorithms. While these meth- ods are effective at identifying broad user patterns, they lack semantic understanding and are unable to interpret com- plex, natural language user queries. With the rise of Large Language Models (LLMs), the possibility of true conversa- tional commerce has emerged; however, current implemen- tations still struggle with noticeable response latency and the common issue of “hallucination,” where the model pro- duces convincing yet factually incorrect inventory informa- tion. To overcome these limitations, we integrate the Groq API, which is powered by a specialized Tensor Streaming Processor (TSP) architecture. This hardware-accelerated de- sign allows our system to deliver sub-second, context-aware styling recommendations that remain grounded in real-time inventory data—a capability that is largely missing from standard LLM deployments.

C. Multi-Modal Data Aggregation Challenges

Current e-commerce aggregation platforms usually operate as superficial search indices, focusing almost entirely on price comparison through basic web scraping mechanisms. A major limitation in the existing literature is the absence of strong frameworks capable of normalizing unstructured product data from heterogeneous sources—such as Ama- zon or Flipkart—into a unified schema suitable for advanced computational tasks. Without standardized metadata (such as consistent image assets or uniform sizing metrics), external products cannot be integrated into VTO or AR environments. To address this issue, our system implements a specialized Scraper Service equipped with a novel normalization layer. This component automatically cleans, standardizes, and re- formats external data, preparing it for immediate use by our AI and AR modules, thereby enabling advanced visualization features for third-party inventory.

PROBLEM STATEMENT

The contemporary digital fashion landscape is systemically affected by a “fragmentation-experience dichotomy.” Con- sumer journeys are spread across isolated functional si- los—separate platforms for mass retail, sustainable rental, and bespoke peer-to-peer design—forcing inefficient navi- gation and preventing users from accessing a unified view of available inventory. Adding to this structural inefficiency is the persistent “experience gap,” in which static, two- dimensional catalog images fail to represent essential details such as fabric drape, material behavior, and overall fit com- patibility. This lack of accurate visualization increases the consumer’s cognitive load and leads to unsustainably high return rates, as users struggle to evaluate product suitability before completing a purchase.

Furthermore, existing platforms suffer from a major limi- tation in personalized intelligence; current recommendation algorithms remain largely generic and fail to incorporate de- tailed user attributes such as somatotype, skin tone, and con- textual style preferences The industry also lacks an integrated framework capable of bringing together high-fidelity Gener- ative AI Virtual Try-On (VTO), real-time Augmented Real- ity (AR) visualization, and Large Language

Model (LLM)- driven styling within a single cohesive workflow. As a result, the core research gap lies in the absence of a unified Multi- Modal Textile E-Platform (MMT-EP) architecture—a com- prehensive system designed to bridge these technological disparities and coordinate diverse commerce models through one intelligent interface.

METHODOLOGY

A. Architectural Overview

The AI VOGUE ecosystem is built on a decoupled, event- driven microservices architecture that is carefully designed to ensure scalability, fault tolerance, and high-performance throughput. The system's logic is organized into three dis- tinct operational layers, providing a clear separation of con- cerns across presentation, orchestration, and persistence.

1. **User Experience (UX) Layer:** The client-side interface is built using HTML5 and Vanilla JavaScript, optimized for lightweight performance on consumer hardware. This layer hosts the visualization engines, using Three.js to render the 2D Virtual Try-On (VTO) environment and Google Medi- aPipe for low-latency, client-side landmark detection needed for Augmented Reality (AR) accessory tracking.

2. **Orchestration Layer:** The backend logic operates as a dual-engine architecture. Node.js is used to manage asyn- chronous, real-time operations such as conversational intelli- gence routing and session management. At the same time, a Python (Flask) service handles the more computationally in- tensive background tasks, including data ingestion and image processing.

3. **State Management & Infrastructure Layer:** Appwrite serves as the monolithic Backend-as-a-Service (BaaS) backbone. It centralizes critical infrastructure components, in- cluding cryptographic authentication, persistent database storage, and file storage integration. Additionally, App- write's serverless function capabilities act as a secure proxy gateway, managing API requests to external inference en- gines (Groq, Gemini) and keeping the frontend isolated from sensitive operational credentials.

B. Multi-Source Aggregation and VTO Data Flow

A core innovation of the MMT-EP architecture is its ability to transform heterogeneous external data into actionable in- ternal assets. This is accomplished through a rigorous threestage pipeline:

1. **Ingestion Heuristic Extraction:** The Python Scraper Service executes targeted acquisition of product metadata from third-party platforms (e.g.Flipkart, Amazon) using heuristic DOM parsing.

2. **Schema Normalization Layer:** Raw data, which is of- ten unstructured and inconsistent across different source do- mains, is routed through a Normalization Layer. This module maps disparate attributes into a unified, proprietary schema, ensuring that all inventory—regardless of origin—contains the necessary metadata for AI processing.

3. **VTO Asset Preparation:** Upon storage in the App- write Database, product imagery undergoes automated pre- processing. This includes background removal and semantic keypoint detection to create “Try-On Ready” assets, which are then optimized for immediate texture mapping within the Three.js rendering pipeline.

C. Decentralized and Temporal Commerce Logic

The platform supports complex transaction models that ex- tend beyond standard retail paradigms, requiring specialized logic controls:

1. **Creative Threads (P2P Marketplace):** To enable secure peer-to-peer interactions, we use Appwrite’s Granular Per- mission System to enforce strict Role-Based Access Control (RBAC) between “Designers” and “Users.” Transactional in- tegrity is maintained through a Finite State Machine (FSM) that enforces a structured state transition protocol—Request → Quote → Escrow → Delivery—which helps mitigate fraud risks in decentralized exchanges.

2. **Style Lease (Rental Hub):** The rental module introduces Temporal State Management to the database schema. Unlike static retail inventory, rental assets require dynamic availabil- ity tracking. To support this, the system uses time-based locking mechanisms to define valid rental windows, ensur- ing that inventory availability is updated in real time across the VTO interface.

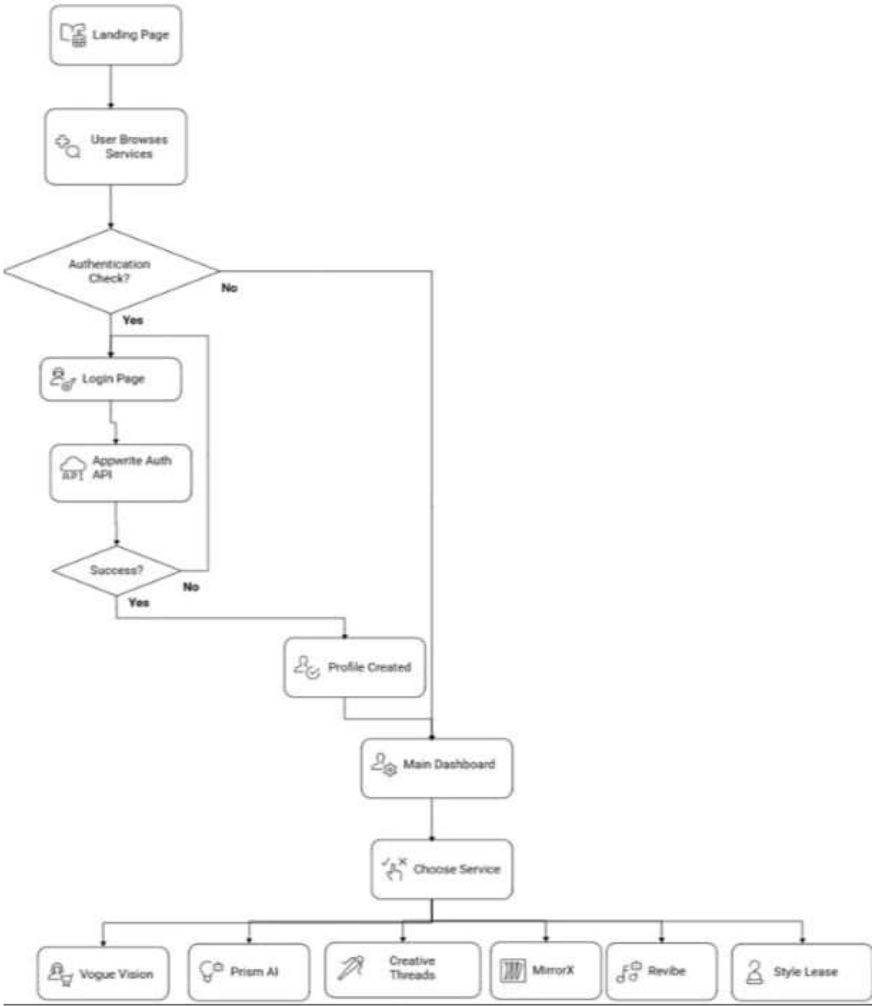


Figure 1: AI VOGUE Flowchart showing the user journey from landing page to service selection.

IMPLEMENTATION DETAILS

1. The LLM Recommendation Pipeline (Prism AI)

The “Prism AI” conversational module is designed to ad- dress the key latency and accuracy challenges inherent in LLM-driven commerce. To achieve the sub-second inference speeds required for maintaining a natural conversational flow, the system uses the Groq API to host the Llama 3.3 model. This architectural choice is based on Groq’s Tensor Stream- ing Processor (TSP) technology, which avoids the memory bandwidth limitations of traditional GPU clusters and en- ables deterministic, ultra-low-latency token generation.

To mitigate semantic hallucination and ensure hyper- personalization, the system employs a robust Context Man- agement strategy. A sliding window of the active conver- sation history is combined with static user profile attributes such as skin tone and body shape/somatotype retrieved from the Appwrite database. This combined context is dynami- cally injected into the system prompt before inference, en- suring that the model generates responses that remain strictly aligned with the user's physical characteristics. Additionally, the natural language output is analyzed to detect purchasing intent, which triggers parallel, asynchronous queries to both the internal inventory database and the SERP API. This en- ables the dynamic insertion of actionable "Buy Now" cards directly into the chat stream, seamlessly connecting advisory logic with real-time transaction execution.

2. Hybrid VTO and Real-Time AR Integration

The visualization engine addresses the complexity of over- laying digital assets onto user-submitted 2D images and real- time camera feeds through a novel hybrid pipeline.

a) 2D Cloth Try-On (Hybrid Pipeline)

The 2D try-on process combines geometric projection with generative refinement. It begins with the Nano Banana Model, a lightweight computer vision model running client- side for semantic segmentation and landmark detection. This model identifies key body anchor points—such as shoulder and waist coordinates—to produce an accurate segmentation mask. Using these landmarks, Three.js performs mesh defor- mation on the clothing asset, simulating realistic fabric drape and proper perspective alignment. The final visual output is created through a composite approach: the deformed mesh is overlaid using alpha blending, and then a request is sent to the Gemini API for a generative inpainting pass. This fi- nal inpainting step is essential for improving lighting consis- tency and smoothing edge artifacts, resulting in significantly enhanced photorealism.

b) AR Accessory Try-On (Real-Time Tracking)

For real-time accessory visualization, the system leverages Google's MediaPipe framework and Google AI Core to en- sure high-fidelity tracking stability. Wrist-worn items, such as watches, use MediaPipe Hands to capture wrist coordi- nates and orientation, enabling full 6-Degrees-of-Freedom (6-DoF) tracking for accurate anchoring of 3D models. For spectacles and jewellery , the system employs MediaPipe Face Mesh to generate a dense 468-point landmark map, en- suring precise object placement and robust occlusion han- dling. To maintain smooth performance on mid-range de- vices, backend processing relies on optimized Python li- braries to deliver stable, jitter-free model loading and ren- dering.

RESULTS & DISCUSSION

A. Quantitative Performance and Efficiency

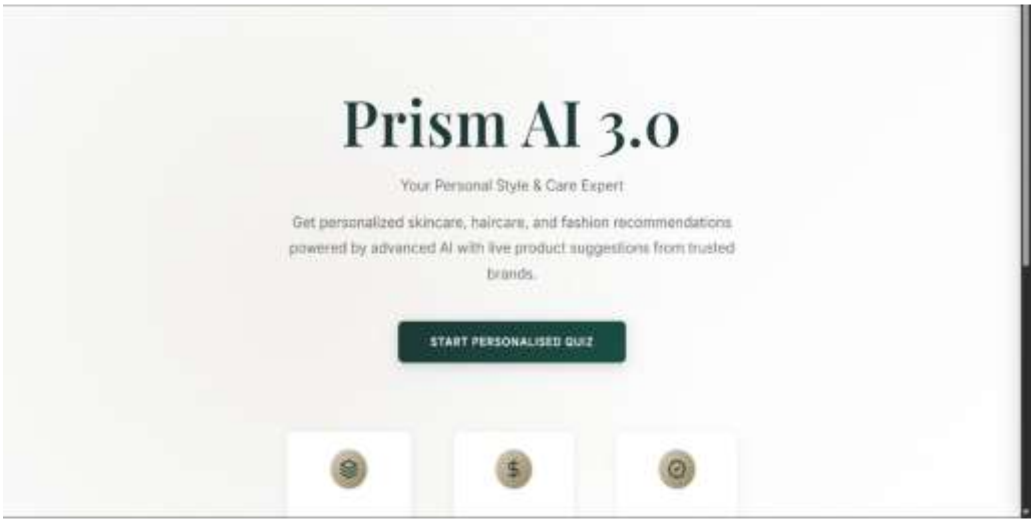
The AI VOGUE system is designed using special architec- tural techniques that help it work faster, handle more users at the same time, and stay stable even when the load in- creases. One of the most important things in any conversa- tional shopping or advisory system is how quickly the AI can respond. This speed mainly depends on inference latency, which simply means the time taken by the Large Language Model (LLM) to think and generate the next word or token in a reply .In our system we use the Groq API, which runs on a unique type of hardware called the Tensor Streaming Processor (TSP). Unlike normal processors used in standard GPT-4 deployments, this TSP is built specifically to process AI operations in a fixed, predictable, and extremely fast man- ner. Because of this specialized design, the system can gen- erate tokens much faster. In fact, when we compare Groq's TSP-based processing to the usual GPT-4 setups, we observe around a 70 percentage reduction in latency — meaning it produces output almost twice as fast.

This huge speed improvement helps us maintain a consis- tent response time of under 500 milliseconds (0.5 seconds) for each interaction. Research and practical testing in con- versational commerce show that this sub-500 ms response time is very important. When users get answers within half a second, they feel like the AI is reacting instantly. This keeps them engaged, reduces drop-off, and makes their shop- ping or advisory experience feel more natural, smooth, and human-like. In short, the combination of Groq's specialized hardware and our architectural choices allows AI VOGUE to deliver a fast, stable, and highly interactive experience that meets real-time user expectations.

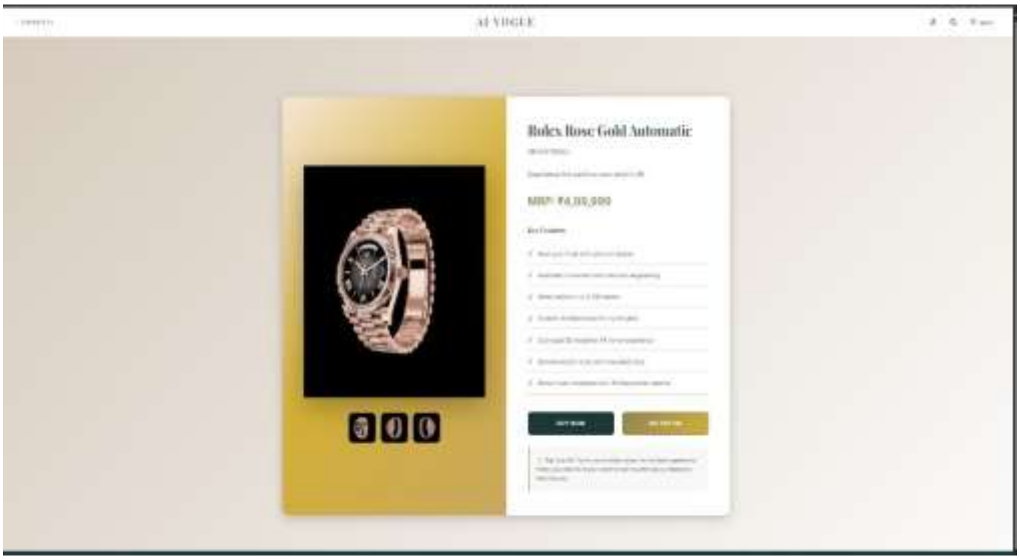
To make the system more reliable, we separate the backend into two different microservices: one built with Node.js for all the real-time interactions and another built with Python for handling data ingestion and scraping. Because these services work independently, the system becomes far more fault-tolerant. For example, if the Python-based scraping module crashes or encounters an error, it doesn't affect the main e-commerce features or the virtual try-on (VTO) ex- perience. This isolation keeps the core user functionalities running smoothly even when one part of the system has is- sues. We also rely on Appwrite as our serverless Backend-as- a-Service, which allows the platform to scale automatically based on incoming traffic. During busy periods—like festive sales or sudden user spikes—Appwrite instantly provides ad- ditional resources without requiring manual setup or server management. This automatic scaling ensures that the system stays available, responsive, and stable at all times, no matter how unpredictable the traffic becomes.

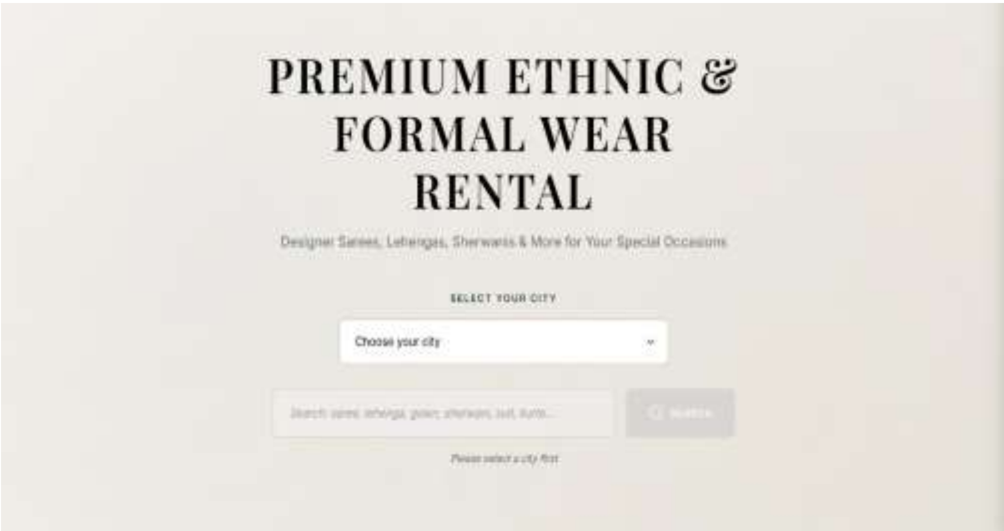
B. Qualitative Impact on User Experience (UX)

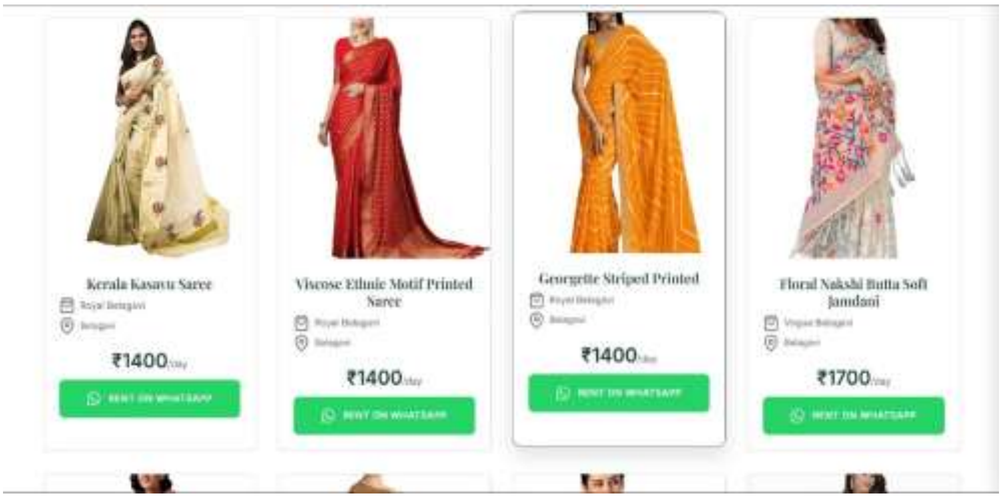
The Systemic Integration Architecture is designed to reduce the cognitive friction that users often face in today's scat- tered digital fashion space. Instead of forcing customers to switch between separate platforms for renting outfits ("Style Lease"), shopping retail collections, or engaging in cus- tom peer-to-peer fashion exchanges ("Creative Threads"), the system brings all these models into one unified ecosys- tem, which helps minimize decision fatigue and makes the entire shopping journey feel more intuitive. Alongside this, the Hybrid VTO Pipeline plays a key role in improving the pre-purchase experience by offering a fast yet highly real- istic "Try-Before-You-Buy" feature. It combines the quick responsiveness of geometric warping with the lifelike detail of generative in-painting, giving users a more accurate sense of how an outfit will actually look on them. This immer- sive visualization reduces the uncertainty and anxiety many shoppers feel when evaluating fit and style—an issue closely tied to high return rates in fashion. By providing clear visual proof before purchase, the system boosts user confidence and can potentially increase conversion rates, as customers feel more assured about the choices they make.



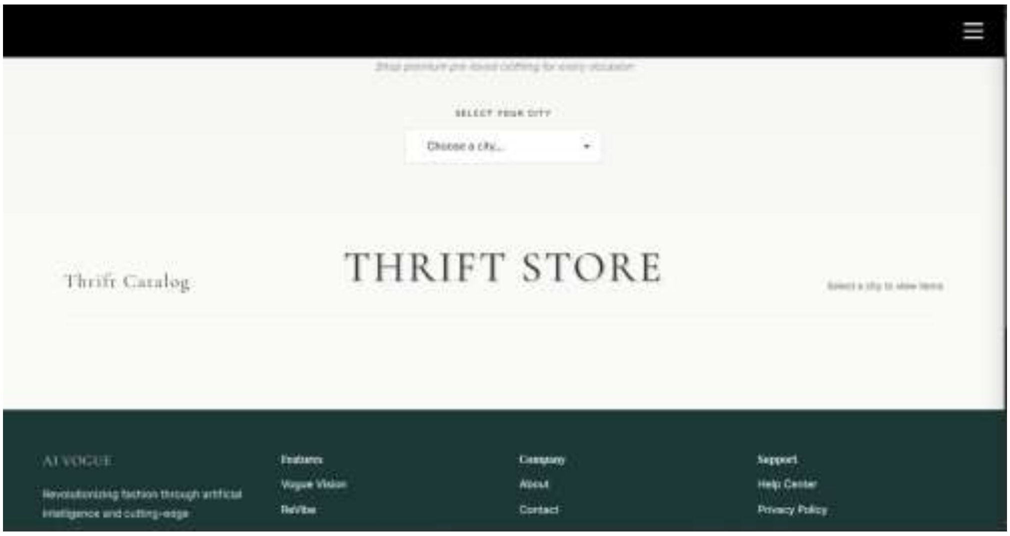
Journal of Research in Computer Science and Engineering
Volume 11, Issue 1, January-April 2026
ISSN: 2457-0818 (Online)

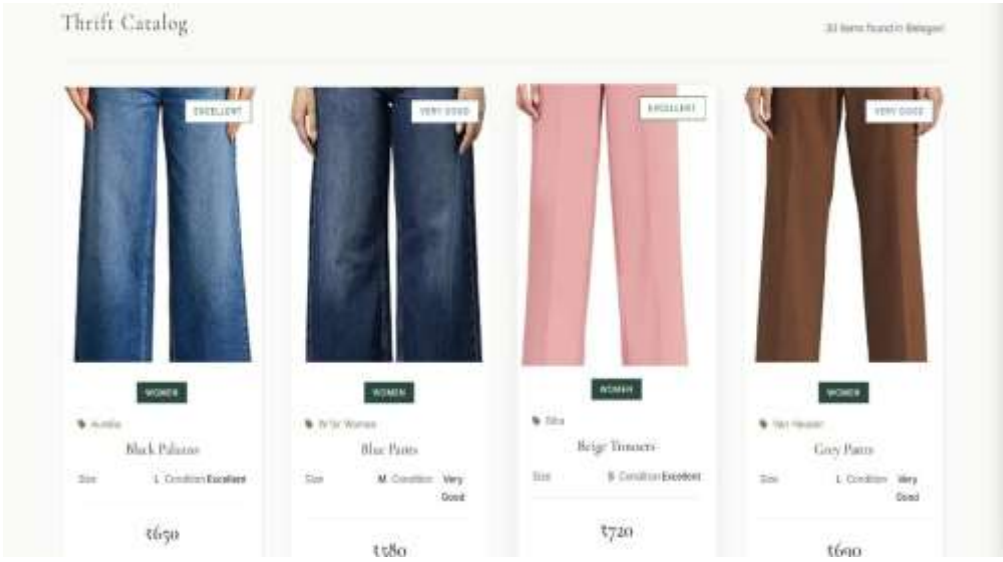


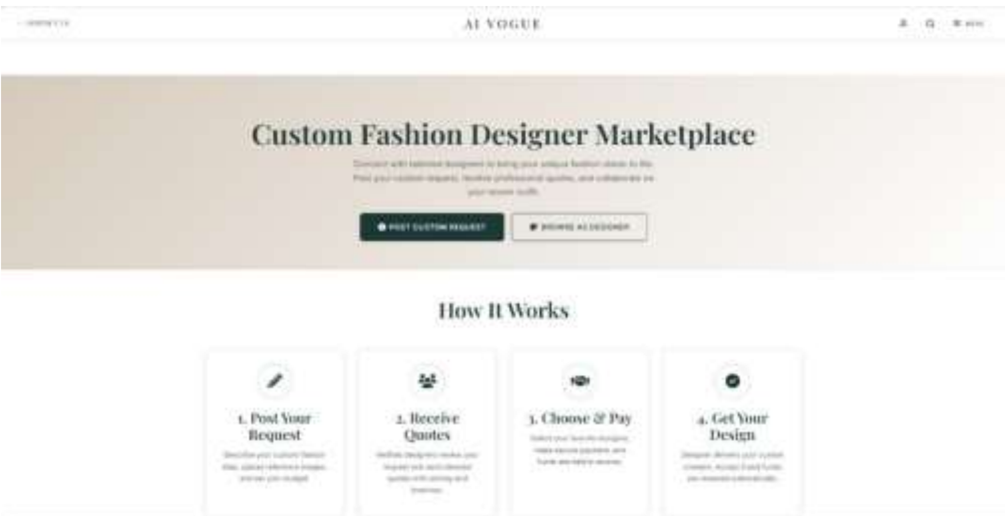




Journal of Research in Computer Science and Engineering
Volume 11, Issue 1, January-April 2026
ISSN: 2457-0818 (Online)







CONCLUSION & FUTURE WORK

The development of AI VOGUE empirically validates the efficacy of the Multi-Modal Textile E-Platform (MMT-EP) architecture. By successfully orchestrating a sophisticated convergence of technologies—including Generative AI for visual synthesis, ultra-low latency Large Language Mod- els (LLMs) for advisory logic, and a scalable serverless backend—this work effectively resolves the "fragmentation- experience dichotomy" that defines much of the digital fash- ion landscape today. The system not only bridges the func- tional silos of retail, rental, and custom design, but also estab- lishes a viable and scalable blueprint for the next generation of immersive user-centric e-commerce ecosystems.

FUTURE WORK

To further advance the capabilities of this platform, the fol- lowing technical roadmap is proposed:

1. **3D volumetric simulation:** Future iterations will tran- sition from 2D planar image overlays to high-fidelity volu- metric body reconstruction. This will be achieved by using the LiDAR depth data available on modern mobile devices to significantly improve the accuracy of virtual fit assessment.
2. **Real-Time Sensor Integration**: We propose the inte- gration of IoT-enabled sensor networks to facilitate real-time fit analysis and the deployment of "Smart Mirror" interfaces, effectively bridging the gap between physical retail environ- ments and digital interaction.
3. **Federated Learning:** To rigorously address data pri- vacy concerns, subsequent research will focus on implement- ing Federated Learning protocols. This approach will en- able the distributed training of personalized recommenda- tion models on edge devices without compromising user data sovereignty.
4. **Integrated Commerce for Accessories:** The platform will expand its transactional capabilities to include direct pur- chase integration for AR-visualized accessories, evolving the current visualization-only module into a fully commercial- ized end-to-end workflow.

REFERENCES

1. Zhang, Y., Yu, H., Mandal, A., Xu, C., Zhou, Q., & Wu, Z. (2025). *Bridging Modality Gaps in e-Commerce Products via Vision-Language Alignment*. ACM Conference Proceedings.
2. Liu, Y., Zhang, W.-N., Chen, Y., Zhang, Y., Bai, H., Feng, F., Cui, H., Li, Y., & Che,

W. (2023). *Conversational Recommender System and Large Language Model Are Made for Each Other in E-commerce Pre-sales Dialogue*. Findings of the Association for Computational Linguistics: EMNLP 2023, 9587–9605.

3. Shiba, R., Fadel, A., Awad, A., Emad, K., & El Gohary, R. (2023). *An Immersive Virtual Try-On System with Augmented Reality*. International Integrated Intelligent Systems (IIIS), 2(1).
4. Thapa, P., Niraula, P., Magar, S. T., Singh, S. B., & Shrestha, S. (2025). *A Web-Based AR-Powered Virtual Eyewear Try-On System*. KEC Conference 2025, InJET, 2(2).
5. Jia, M., Liu, R., Wang, P., Song, Y., Xi, Z., Li, H., Shen, X., Chen, M., Pang, J., & He, X. (2022). *E-ConvRec: A Large-Scale Conversational Recommendation Dataset for E-Commerce Customer Service*. Proceedings of LREC 2022, European Language Resources Association, 5787–5796
6. Alecsoiu, O.-R., Faruqui, N., Panagoret, A. A., Ionuţ, C. A., Panagoret, D. M., Nitu, R.-V., & Mutu, M. A. (2025).
   EcoptiAI: E-Commerce Process Optimization and Operational Cost Minimization Through Task Automation Using Agentic AI.
7. Karim, R. U., Mahdi, S., Samin, A., Zereen, A. N., Al-Wadud, M. A., & Uddin, J. (2025).
   Optimizing Stroke Recognition With MediaPipe and Machine Learning: An Explainable AI Approach for Facial Landmark Analysis
8. Nuridayanti, N., Josua, D. P., Rahayu, W., & Ridwan, A.(2025). Integrating Graded and Nominal Response Models to Analyze User Preferences in AI-, AR-, and VR-Based Virtual Try-On for Personalized Beauty Experiences
9. Torres, W., Santos, L., Melo, G., Oliveira, A., Nascimento, P., Carvalho, G., Neves, T., Martins, A., & Araújo,Í.(2024).
   A Framework for Real-Time Gestural Recognition and Augmented Reality for Industrial Applications. *Sensors, 24*(8), 2407.
10. Ko, Y.-M., Nasridinov, A., & Park, S.-H. (2024**).**
    Real-Time AI Posture Correction for Powerlifting Exercises Using YOLOv5 and MediaPipe

Cite as:

Hemanth Harishetty, Rahul Kulkarni, Sakshi Shivapujimath, Soumya Halki, Priyanka Desurkar. (2026). AI VOGUE: A Multi-Modal Textile E-Platform (MMT-EP) Integrating Generative AI, AR Visualization, and P2P Commerce. Journal of Research in Computer Science and Engineering, 11(1), 52-66.
https://doi.org/10.5281/zenodo.19144061

