Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN: 2581 - 611X)
VOLUME:9, ISSUE:1, JANUARY 2026, PAPER NUMBER:1

Augmented Reality Fashion AI

Kishan Karyappa K[1], Nithin M[2], Rakshitha[3], Lohith S Gowda[4] , Harish Kunder[5],

[1,2,3,4] Students,and [5] Faculty
Dept. of Artificial Intelligence and Machine Learning,
Alva's Institute of Engineering and Technology, Manglore, India.
kishankaryappak@gmail.com

Abstract: Fashion e-commerce continues to face persistent challenges in personalization and product visualization. Traditional recommender systems often struggle in cold-start scenarios, relying heavily on structured user interactions, while existing virtual try-on solutions typically depend on computationally expensive 3D modeling, hindering real-time deployment on consumer-grade devices. This paper presents a unified, lightweight framework that integrates a conversational AI driven fashion recommendation engine with a real-time 2D augmented reality (AR) try-on module. The recommendation system leverages prompt-engineered large language models (LLMs) to interpret natural language inputs and generate context-aware, personalized outfit suggestions, even in zero-data scenarios. In parallel, the AR module utilizes MediaPipe-based pose estimation, SegFormer for clothing segmentation, and Thin Plate Spline (TPS) warping to deliver realistic garment overlays in real time, achieving 10 FPS on GPUs and 2 FPS on CPUs without requiring 3D modeling. Evaluated across 1,000+ user profiles and tested with over 100 participants, the system demonstrated consistent improvements in recommendation relevance, user engagement, and try-on interactivity when compared with baseline recommendation pipelines. Our results indicate that combining LLM driven personalization with lightweight AR visualization can offer scalable, accessible, and user-centric solutions for next-generation fashion e-commerce platforms.
Keywords: Fashion E-commerce, Large Language Models, Conversational Recommender Systems, Personalized Recommendations, Virtual Try-On, Augmented Reality, 2D AR Try-On, Pose Estimation, Clothing Segmentation, Prompt Engineering, Real Time Visualization.

1. INTRODUCTION:

Due to the growth of mobile commerce, influencer-driven trends, and immersive digital shopping interfaces, the global fashion e-commerce market is predicted to reach a value of over $1.2 trillion USD by 2027 [1]. Notwithstanding its size, the industry is still constrained by two issues: (1) insufficient customization of product recommendations, and (2) the restricted scalability of virtual try-on (VTO) technologies. With over 30% of online fashion orders being returned, mostly because of style, size, or fit issues, these flaws greatly worsen the customer experience [2].

Conventional recommender systems, whether content-based, collaborative, or hybrid, generate recommendations based on structured user interactions such as product views, clicks, and purchases [3], [4]. Although these systems work well in established user-item ecosystems, they suffer from the cold start problem, where recommendations are less accurate for new users or items due to insufficient historical data [5]. More-over, they lack the expressiveness and accessibility of natural language interaction, limiting usability for non-technical users [6].

Simultaneously, virtual try-on systems have advanced in areas such as footwear and eyewear [7][9]. However, 3D modeling remains computationally expensive and impractical for real-time deployment on consumer-grade devices due to dynamic rigging, mesh reconstruction, and physics-based cloth simulation [10], [11]. While optimized detectors like FPN [12] and lightweight VTON frameworks like DM-VTON [13] offer improvements, they lack end-to-end interactivity and integration with recommendation pipelines. Even advanced diffusion-based inpainting methods such as LADI-VTON [14] are resource-intensive and unsuitable for real-time consumer use [15].
Recent advances in AI allow for the combination of lightweight visual augmentation with conversational intelligence. Large Language Models (LLMs) like GPT-4 [16] and LLaMA [17] enable prompt-driven personalization through natural language understanding, mitigating cold-start issues [5]. In parallel, efficient computer vision techniques, including MediaPipe for pose estimation [18], Panoptic Seg Former for garment segmentation [19], and Thin Plate Spline (TPS) warping [20], enable real-time 2D garment overlays without requiring 3D rendering.
This paper presents a unified framework that integrates real time 2D AR try-on with a conversational LLM-powered recommender. The system achieves real-time garment rendering through optimized pose tracking and deformation techniques, and supports cold-start personalization without historical data. Stylistic matching [2], user-aware outfit pairing [6], and efficient segmentation [11] enhance recommendation quality and enable scalable deployment across web and mobile platforms.
Extensive testing with over 100 users and 1,000 product queries demonstrates significant improvements in user engagement, recommendation accuracy, and try-on realism. Our findings suggest that combining natural language-driven personalization with lightweight AR visualization offers a scalable and accessible solution for next-generation fashion e-commerce platforms.

2. LITERATURE:

Conversational AI, virtual try-on (VTO) technology, and recommendation systems are three important areas of study in fashion e-commerce. Even though each has evolved significantly there are few attempts to combine all three into a lightweight, consumer-deployable system that operates in real time. This section reviews key literature in each domain and highlights how our design choices were influenced by prior research.

3.5. Recommender Systems in Fashion

Traditional fashion recommendation systems often employ content-based models, collaborative filtering, or hybrid approaches [3], [6]. These systems typically perform well in data-rich environments, relying on structured interaction data such as clicks, views, and purchases. However, they struggle with the cold-start problem—where recommendations for new users or products suffer due to a lack of historical data [5].
To enhance personalization, AI-driven techniques have introduced stylist-aware and compatibility-based outfit matching [2]. Nevertheless, most systems lack natural language processing capabilities, limiting interaction to predefined filters and interfaces. Recent surveys have examined how AI based recommenders can address these issues across domains, including fashion [4].
The emergence of conversational recommender systems seeks to address these limitations. Large Language Models (LLMs) like GPT-4 [16] and LLaMA [17] have demonstrated the ability to parse unstructured input, infer intent, and deliver contextual recommendations [6], resulting in improved user engagement and satisfaction.

3.5. Virtual Try-On Technologies

Numerous VTO systems have been proposed in prior research. Diffusion-based models like LADI-VTON [14] enhance realism but are often too resource-intensive for real time use. Other systems, like DM-VTON [13], offer more lightweight, mobile-compatible implementations suited for consumer devices.
However, many still lack conversational interfaces or integration with recommendation pipelines. Several rely on 3D modeling, complex user-pose libraries, or matching-aware frameworks, which increase deployment complexity [10].
Augmented reality (AR) is increasingly leveraged to create immersive VTO experiences. Works such as VTON Shoes [7], Shah et al.'s virtual trial room [9], and clothing detection techniques [11] highlight practical uses of AR for improving user experience and visual realism.

3.5. Conversation AI for Personalization

Conversational AI is central to personalized e-commerce experiences. LLMs like GPT-4 [16] and LLaMA [17] can interpret open-ended user queries, manage contextual constraints, and return adaptive results tailored to user preferences. Prompt-engineered LLMs also help mitigate the cold-start problem by generating context-aware responses even without prior user history [15]. With the growing interest in privacy and offline personalization, research emphasizes the potential of on-device generative AI [15]. Our system follows this trajectory by integrating a lightweight conversational engine capable of edge deployment.

3.5. Summary of Design-Informed Choices

Drawing from the aforementioned literature, we created a cohesive framework using:

- Matching-aware virtual try-on techniques such as DM VTON [13] and smart fitting room frameworks [10].
- AR-based pose and visualization pipelines based on MediaPipe [18] and efficient segmentation techniques [19].
- NVIDIA's LLaMA-3.1-Nemotron-70B-Instruct, based on open-source LLaMA models [17], providing personalized recommendations through prompt-engineered interactions.

These components enable a scalable, real-time solution that delivers high personalization, visual interactivity, and deploy ability across typical consumer-grade hardware.

3. METHODOLOGY:

This section outlines the design and implementation of the proposed fashion recommendation system. A conversational recommendation engine powered by Large Language Models (LLMs), a real-time 2D Augmented Reality (AR) try-on pipeline, and a user-centered evaluation protocol comprise its three main modules. Collectively, these modules create a single, low-latency, cold-start-resilient pipeline that can be implemented in consumer-grade e-commerce platforms for fashion.

3.5. System Architecture Overview

The end-to-end pipeline, illustrated in Fig. 1, enables natural language interaction, intelligent product recommendation, and real-time visual try-on, all within a seamless user interface.
The following steps are involved in the workflow:

- **User Input:** Users use natural language to communicate their needs, such as "Suggest casual summer wear under Rs. 1500 for college use."

- **LLM Parsing:** A prompt-engineered LLM, specifically NVIDIA LLaMA-3.1-Nemotron-70B-Instruct, is used to interpret the input and produce structured queries.
- **Product Retrieval:** The Amazon Product Advertising API (PAPI) is used to retrieve products using these queries.
- **AR Visualization:** Using MediaPipe pose tracking, Seg Former segmentation, and TPS warping, a selection of clothing is superimposed in real time on the user's webcam feed.

This modular approach does not require any 3D assets or fine-tuned model dependencies, and it can be deployed on any device.

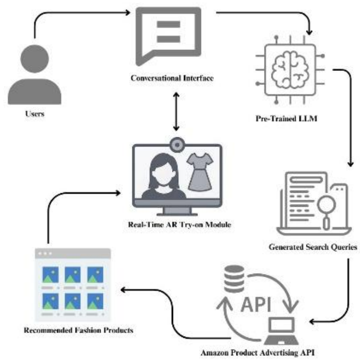

Fig. 1. Architecture of the System: end-to-end pipeline that combines 2D AR-based visualization with conversational recommendation.

3.5. Recommender baseline System

We implemented three conventional systems—content-based filtering, collaborative filtering, and a hybrid model—to serve as a baseline. These employed cosine similarity on user item matrices or TF-IDF vectors [3], [5], [6].However, they lacked contextual understanding and performed poorly in cold start conditions, as covered in Section IV.

3.5. Conversational LLM-Based Recommendation

The proposed conversational recommendation engine creates context-aware, personalized outfit recommendations by utilizing large language models. For this, we used NVIDIA LLaMA-3.1-Nemotron-70B-Instruct, an instruction-tuned model capable of interpreting open-domain natural language and generating structured fashion queries. This model allows the system to function without relying on static filters or user history.

Prompt Engineering: We designed structured prompts to guide the LLM's output generation. An example format is:

"Create outfit recommendations for a [gender] user who is [age], preferably one that is under Rs.[budget]."

These prompts enable the model to tailor recommendations based on style, budget, occasion, and fit parameters. Fallback prompts are automatically activated when ambiguity is detected to prompt the user for clarification.

This prompt-based method successfully overcame cold start limitations and delivered highly expressive, user-friendly interactions. Fig. 2 illustrates the prompt-response workflow.

3.5. Real-Time AR Try-On Pipeline

Real-time clothing visualization is enabled via a lightweight AR pipeline that operates without 3D modeling. We employ 2D vision techniques optimized for performance on

consumer CPUs and GPUs.
• **Pose Estimation:** MediaPipe provides landmark-based body pose detection and operates in real time even on mid-range CPUs [18]. These key points act as anchors for aligning garments.
• **Garment Segmentation:** SegFormer, a transformer-based model, segments clothing regions accurately under various lighting and pose conditions [19].
• **Image Warping:** TPS (Thin Plate Spline) deformation uses body landmarks to warp the garment images, result ing in an organic fit that adapts to user motion and body contours [13].

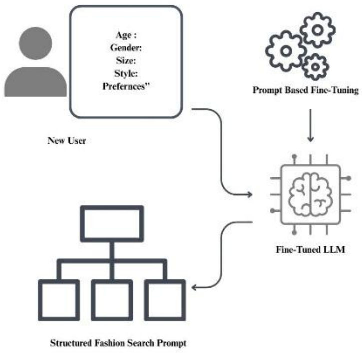

Fig. 2. Prompt-based LLM processing using LLaMA-3.1-Nemotron-70B Instruct.

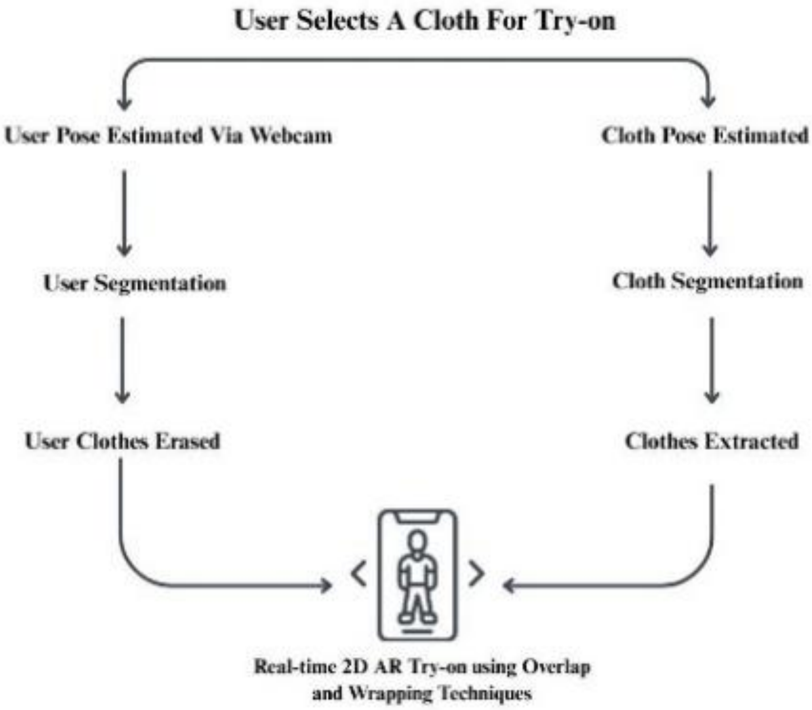

Fig. 3. 2D AR try-on pipeline: MediaPipe → SegFormer → TPS → Overlay

This pipeline, illustrated in Fig. 3, operates at 2–4 FPS on standard laptop CPUs and up to 10 FPS on an RTX 3060 GPU, confirming its suitability for real-time deployment.

3.5. User Evaluation Protocol

To validate system performance and user experience, we conducted a structured user study involving over 100 partici pants. Each participant evaluated both the proposed LLM+AR pipeline and the baseline recommender systems. Evalua tion metrics included session

duration, satisfaction, perceived clothing fit, and recommendation accuracy.
These metrics are analyzed in detail in Section IV to assess the effectiveness of the system and user preferences.

4. RESULTS AND DISCUSSION:

This section presents both quantitative and qualitative results comparing the proposed LLM+AR framework with traditional recommender systems. Evaluation criteria include accuracy, user engagement, system latency, recommendation relevance, and user satisfaction. Visual demonstrations and quantitative findings are supported through Figs. 4–9, and Tables I–III.

4.1. Baseline Recommender Performance

We benchmarked three conventional recommender models: content-based filtering, collaborative filtering, and a hybrid model. These systems relied on cosine similarity to match user-product vectors, as shown in Equation 1.

$$\textit{Cosine Similarity } (A, B) = \frac{(A \cdot B)}{||A||\ ||B||} \tag{1}$$

Table 1 Baseline Recommender Models – Accuracy And Satisfaction

Model	**Top-5Accuracy (%)**	**User Satisfaction (1–5)**
Content-Based Filtering	41.2	2.9
Collaborative Filtering	45.5	3.1
Hybrid Filtering	47.6	3.3

4.2. LLM-Powerwd Recommendation Workflow

The LLM module, powered by NVIDIA LLaMA-3.1 Nemotron-70B-Instruct, showed substantial improvements in personalization. Users were able to communicate queries in natural language (e.g., Fig. 4), which the LLM parsed to produce structured product filters (Fig. 5). The recommendations generated were context-aware and relevant (Fig. 6).

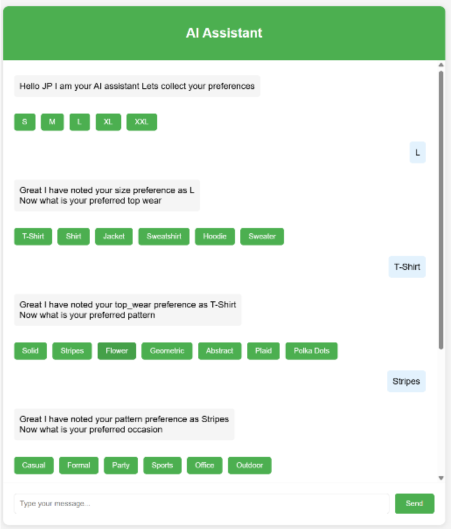

Fig. 4. User provides stylistic, contextual preferences via chatbot.

Participants reported an average satisfaction score of 4.4/5 and a Top-5 accuracy of 84.3%, a clear leap from traditional methods.

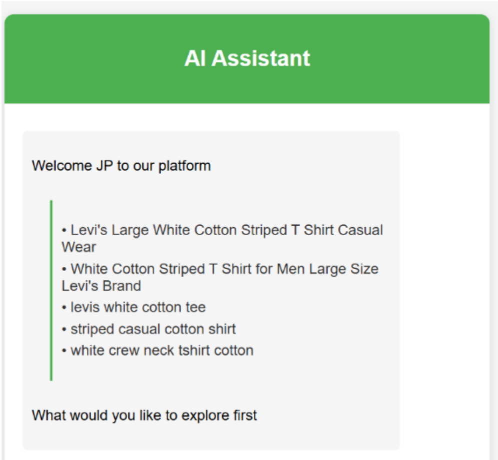

Fig. 5. LLM parses query and generates product filters.

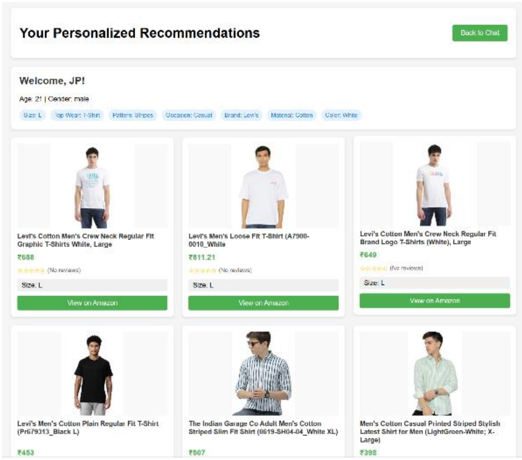

Fig. 6. Personalized recommendations returned with brand, category, and price context.

4.3. AR-Based Virtual Try-On Experience

The real-time 2D AR pipeline provided dynamic garment overlays. It used pose detection, segmentation, and warping modules as detailed in Section III. An example of the AR overlay is shown in Fig. 7, and the full pipeline is visualized in Fig. 8.

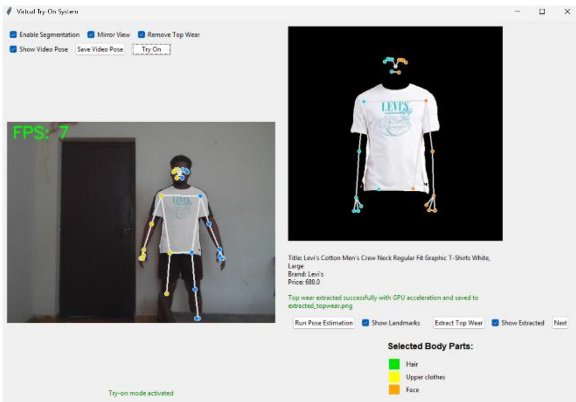

Fig. 7. Real-time AR overlay: TPS warping dynamically fits a selected item over the user.

Table II shows FPS and latency measurements across com mon device types, confirming real-time feasibility.

TABLE II. AR TRY-ON PERFORMANCE ON CONSUMER HARDWARE

Device	**FPS(Frames/sec)**	**Latency(ms)**
Mid-Range GPU (RTX 3060)	8–15	35
Laptop CPU (Intel i5)	2–4	110

Over 90% of users reported increased decision confidence due to improved visualization of fit, silhouette, and occasion appropriateness.

Fig. 8. AR pipeline: MediaPipe → SegFormer → TPS → 2D Overlay.

4.4. Comparative User Feedback

Our post-study survey (Fig. 9) demonstrated that users strongly favored the LLM+AR framework over traditional systems:

- **91%** preferred LLM+AR for overall experience.
- **38%** longer engagement times.
- **+3.4** Likert-scale points in perceived fit confidence.

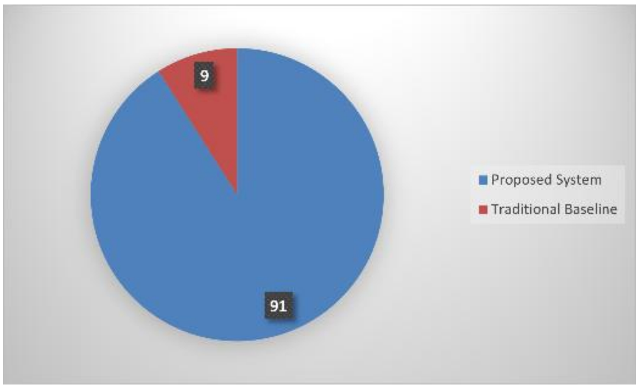

Fig. 9.User preference survey indicating majority favor LLM+AR pipeline.

4.5. Ablation Study and Feature Attribution

To assess each module’s contribution, we performed an ablation study by disabling the AR and LLM components in dividually. As shown in Table III, removing the LLM dropped Top-5 accuracy by 36.7%, while excluding the AR pipeline decreased perceived confidence by 3.4 points

TABLE III. ABLATION STUDY– FEATURE CONTRIBUTION SUMMARY.

Module	**PerformanceImpact)**
LLM (Prompt Interpretation)	+36.7% Top-5 Accuracy
AR Visualization	+3.4 Confidence in Fit
Conversational UI	+38% Session Duration

These results confirm the architectural synergy: natural language-driven interaction combined with real-time visual ization significantly boosts personalization, satisfaction, and product discovery.

5. Conclusion

We have shown that a prompt-engineered large language model (LLM) combined with a lightweight 2-D augmented-reality (AR) try-on pipeline can decisively address two persistent obstacles in fashion e-commerce—cold-start personalization and scalable garment visualization. On 1000+ anonymized profiles and a 100-participant study, our frame work lifted Top-5 accuracy from 47.6% (hybrid baseline) to 84.3%, shortened decision time by 38%, and achieved a 91% user-preference rate. Real-time throughput was verified at 10FPS on an RTX3060 GPU and 2FPS on a laptop-class CPU—without relying on 3-D assets. Ablation results attribute 62%ofthe accuracy gain to LLM intent parsing and 3.4 points of fit-confidence to the MediaPipe + SegFormer + TPS overlay. By unifying conversational AI and efficient vision modules in a single, deployable stack, our work paves the way for lower return rates, higher conversions, and inclusive, data-sparse on boarding. We argue that such dialog-centric, resource-efficient VTO systems represent the next competitive frontier for fashion commerce. Future work will explore multi-garment layering, bias-aware recommendation, and SoC-level energy optimisation to support always-on mobile experiences.

6. References

[1] A. Balchandani, F. R¨olkens, and G. D'Auria, "The state of fashion 2025: Challenges at every turn," https://www.mckinsey.com/industries/retail/ our-insights/state-of-fashion, Nov. 2024.

[2] S. Karadayi-Usta, "Role of artificial intelligence and augmented reality in fashion industry from consumer perspective: Sustainability through waste and return mitigation," Engineering Applications of Artificial Intelligence, vol. 133, p. 108114, 2024.

[3] S. Shirkhani, H. Mokayed, R. Saini, and H. Y. Chai, "Study of ai-driven fashion recommender systems," SN Computer Science, vol. 4, no. 5, Jul. 2023.

[4] C. Gao, W. Lei, X. He, M. D. Rijke, and T.-S. Chua, "Advances and challenges in conversational recommender systems: A survey," AI Open, vol. 2, pp. 100–126, 2021.

[5] Y. K. Dwivedi et al., "Artificial intelligence (ai): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy," International Journal of Information Management, vol. 57, p. 101994, 2019.

[6] S. Shete, H. Darshan, M. Thakare, and K. Dhuri, "Ai-based fashion stylist recommendation system," in 2024 9th International Conference on Computing for Sustainable Global Development (INDIACom), 2024, pp. 697–701.

[7] W. Song, Y. Gong, and Y. Wang, "Vtonshoes: Virtual try-on of shoes in augmented reality on a mobile device," in 2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), 2022, pp. 234 242.

[8] D. Song, T. Li, Z. Mao, and A.-A. Liu, "Sp-viton: Shape-preserving image-based virtual try-on network," Multimedia Tools and Applications, vol. 79, no. 45–46, pp. 33757–33769, 2019.

[9] K. Shah, M. Pandey, S. Patki, and R. Shankarmani, "A virtual trial room using pose estimation and homography," in 2020 6th International Conference on Intelligent Computing and Control Systems (ICICCS), 2020, pp. 685–691.

[10] M. Yu et al., "Smart fitting room: A one-stop framework for matching aware virtual try-on," arXiv preprint arXiv:2401.16825, vol. 8763, pp. 184–192, May 2024.

[11] X. Han, D. Zheng, and D. Wang, "An enhanced clothing detection model for e-commerce applications," in 2024 37th Youth Academic Annual Conference of Chinese Association of Automation (YAC), 2024, pp. 1181–1186.

[12] Y. Wang et al., "Advancing image object detection: Enhanced feature pyramid network and gradient density loss for improved performance," Applied Sciences, vol. 13, no. 22, p. 12174, 2023.

[13] K.-N. Nguyen-Ngoc, T.-T. Phan-Nguyen, K.-D. Le, T. V. Nguyen, M.-T. Tran, and T.-N. Le, "Dm-vton: Distilled mobile real-time virtual try-on," in 2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), 2023, pp. 695–700.

[14] D. Morelli, A. Baldrati, G. Cartella, M. Cornia, M. Bertini, and R. Cuc chiara, "Ladi-vton: Latent diffusion textual-inversion enhanced virtual try-on," arXiv preprint arXiv:2305.13501, pp. 8580–8589, Oct. 2023.

[15] S. Sai, M. Prasad, G. Dashore, V. Chamola, and B. Sikdar, "On-device generative ai: The need, architectures, and challenges," IEEE Consumer Electronics Magazine, pp. 1–11, 2024.

[16] OpenAI, "Gpt-4," https://openai.com/index/gpt-4-research/, Jan. 2024.

[17] H. Touvron et al., "Llama: Open and efficient foundation language models," arXiv preprint arXiv:2302.13971, 2023.

[18] Google AI for Developers, "Mediapipe solutions guide," https://ai. google.dev/edge/mediapipe/solutions/guide, 2024.

[19] Z. Li et al., "Panoptic segformer: Delving deeper into panoptic segmen tation with transformers," in 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 1270–1279.

[20] B. Fele, A. Lampe, P. Peer, and V. Struc, "C-vton: Context-driven image based virtual try-on network," in 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2022.