Custom Cart – Virtual try-on in e-commerce platforms using generative AI

Jesvin Jelson Thachil
Department of Computer Science and Engineering
Amal Jyothi College of Engineering
Kanjirapally,Kottayam
jesvinjelsonthachil2026@cs.ajce.in

Kesiya Rachel Johns
Department of Computer Science and Engineering
Amal Jyothi College of Engineering
Kanjirapally,Kottayam
kesiyaracheljohns2026@cs.ajce.in

Mehak Riyas
Department of Computer Science and Engineering
Amal Jyothi College of Engineering
Kanjirapally,Kottayam
mehakriyas2026@cs.ajce.in

Ken Jacob Zachariah
Department of Computer Science and Engineering
Amal Jyothi College of Engineering
Kanjirapally,Kottayam
kenjacobzechariah2026@cs.ajce.in

Neenu R
Assistant Professor
Department of Computer Science and Engineering
Amal Jyothi College of Engineering
Kanjirapally,Kottayam
rneenu@amaljyothi.ac.in

***Abstract*—Custom Cart is a virtual try-on e-commerce system that uses generative AI to Allow users to virtually try-on different clothing. The system enables customers to upload their photo and visualize themselves wearing clothing of their selection, simulating an in-store fitting room. This feature reduces uncertainty about their purchases, minimizing return rates. The platform leverages COMFYUI, a Python-based framework, that works by using different nodes to process images and generate realistic try-on results. By integrating large language models for masking and creating segments in the user's image and image-generation AI, Custom Cart will improve personalization, improve user engagement, and enhance the overall e-commerce experience. The combination of fast processing speeds, with the use of high-speed servers and advanced AI models makes it a promising solution for the future of online fashion retail.**

Keywords— Generative Artificial Intelligence, Large Language Models, Virtual Try-On, Computer Vision, E-commerce, User Experience Design, Sustainable Technology, Artificial Intelligence, E-commerce Application.

I. INTRODUCTION

Custom Cart is an innovative e-commerce system that uses image-generation AI to enhance the online shopping experience. It addresses a problem of consumers returning items because they are unhappy with the way clothing looks on them [1]. The system offers a realistic visual preview by enabling customers to virtually try on clothing by uploading a photo of themselves, which boosts user confidence and lowers returns. Options for personalized gifts for different occasions enhance the user experience even more. COMFYUI, a Python-based tool is used for the image generation using large data models and is integrated into the website which precisely places the selected clothing over the user's photo and guarantees high-quality visual results. COMFY UI needs to be run-on high-performance servers for fast and accurate image processing.

A. User Convenience

Users can easily access the platform through any web browser without requiring additional software or installations. The virtual try-on feature is straightforward—a system where users can simply upload a photo clicking on a try on button of a product and it will redirect them to a pop up where the image will be processed and the user can put clothing on themselves.

B. Quick and Seamless Interaction

This system provides a user-friendly experience with smooth navigation. While the virtual try-on feature offers accurate and realistic previews, faster processing speeds are only achievable with high-performance servers. The platform ensures users can still interact easily, with optimized processes balancing speed and resource usage. The user's image is uploaded to a fast API python server which then sends that image to COMFY UI which is run on another separate server or a different port. After the user's image is processed by a custom workflow designed for swapping of clothes, created using nodes in COMFY UI. The image is fetched back to the main server and displayed.

In the COMFY UI workflow (fig 1.1), it uses two image data models to create a mask around the user's clothes and replace it with the input clothes given in the website.

C. Cross-Device Compatibility

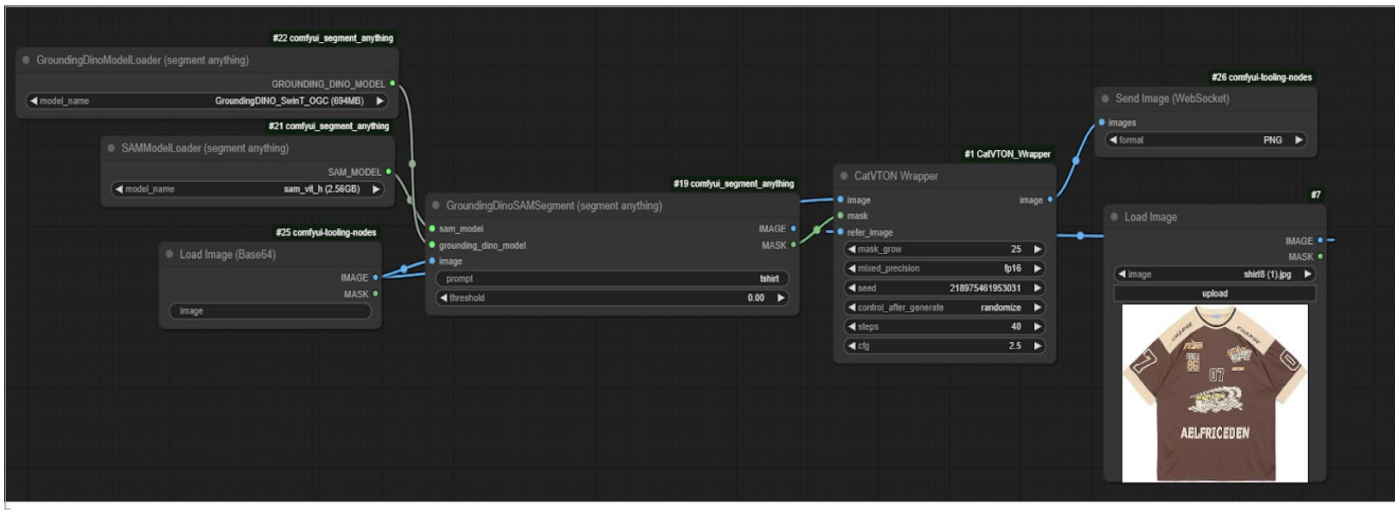

Fig 1.1

Custom Cart is designed to be accessible across multiple devices, including desktops and smartphones. The platform adapts to different screen sizes, providing a consistent and optimized experience for all users.

In smartphones the users can either choose to upload an existing photo of themselves, or take a picture using their camera.

D. Precision and Accuracy

The precision and accuracy of the clothing depends on a good lighting environment and quality of the image. The resultant image will be accurate in perfect environments.

E. COMFY UI

Comfy UI is an open-source, node-based user interface used for generating images from text prompts and for using large image data models like Stable Diffusion. Created by comfy anonymous and released in January 2023, it is a flexible platform for AI-driven image synthesis.

Comfy UI is the main backend server used for the clothing swapping done in the e-commerce platform while performing Virtual Try-On. This is achieved by making a custom API for fetching and uploading images to Comfy UI. The bridge between the web page and the Comfy UI server is a fast API python server.

II. SYSTEM DESIGN

A.Frontend Architecture

The products are first shown as a catalog where the user can select which product to try on as shown in fig 1.2.

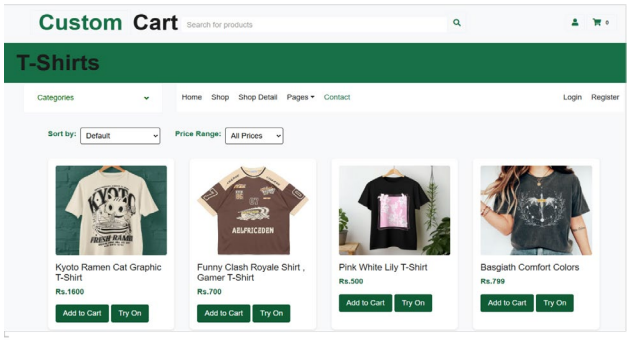

Fig 1.2

After clicking on try on, the user will be redirected to a pop up as shown in fig 1.3 where they can upload their image. The uploaded image is then sent to the fast API server. The Comfy UI server will receive the prompt and start processing the image.

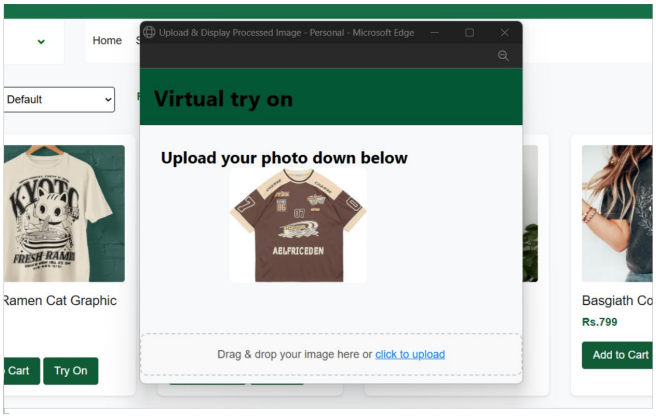

Fig 1.3

The user can upload an existing photo of themselves (or take a picture on their mobile device).

After image processing the image will be displayed in the pop up as shown in fig 1.4.

Fig 1.4

B.Backend COMFY UI architecture

- Each product image needs to have its own COMFY UI workflow. The workflow is a .json file which contains the process through which COMFYUI runs.
- The custom COMFY UI workflow used in virtual try on is shown in fig1.1.

C. Data models used

The sam_vit_h model is a variant of the Segment Anything Model (SAM) [3] that uses the Vision Transformer Huge (ViT-H) backbone. This configuration comprises approximately 636 million parameters, offering the highest quality outputs among SAM's available encoders, though it operates at a slower speed compared to the ViT-B (Base) and ViT-L (Large) versions. This data model is perfect for accurately segmenting users' apparels.

The GroundingDINO_SwinT_OGC [3] model is a variant of the Grounding DINO framework, which integrates the DINO (DETR with Improved DeNoising Anchor Boxes) object detection model with new pre-training techniques. This fusion enables the detection of arbitrary objects based on user inputs, such as category names or referring expressions. The SwinT component refers to the Swin Transformer Tiny backbone utilized in this model, which processes images to extract hierarchical feature representations.

D. Cloth swapped image production process

- The Load Image (Base64) node in ComfyUI serves to decode Base64-encoded image data into an image format usable in the workflow. It supports smooth image input handling, thus being beneficial for web applications like custom cart where images are commonly sent in Base64 format. The node provides efficient processing by decoding encoded images into a standard format for additional AI-based processing. It is implemented in virtual try-on systems that allow users to upload pictures straight from their phones for real-time processing.
- Each product has its own workflow which is accessed by the product id. The Load Image node loads the products reference clothing image.
- SAMModelLoader loads the sam_vit_h model (2.5GB) for image segmentation.
- GroundingDinoModelLoader loads another data model GroundingDINO_SwinT_OGC for object detection and segmentation. This helps to mask the user's clothing.
- GroundingDinoSAMSegment processes the loaded models to segment and detect the clothing area from the input image. The "tshirt/top" prompt assists the model in targeting the intended clothing area, which leads to accurate segmentation. A threshold (0.30) sets segmentation accuracy by regulating the area of region selection. This node generates the required mask for the CatVTON wrapper node, facilitating realistic virtual try-on.
- The CatVTON Wrapper Node subsequently places the reference clothing image over the segmented area of the user's uploaded image by blending the base image, mask, and reference image. The mask grow (25) parameter increases the size of the segmented area slightly so that it fits smoothly and naturally without jarring edges or distortions (but can occur with low light environments). The node also makes use of mixed precision (fp16), which balances computational efficiency with high-quality output, maintaining a compromise between computational speed and visual quality. To add variations and avoid repetition in the outputs, controlled randomization is implemented through the seed parameter to produce slightly different but realistic variations for every processed image. Also, the steps (40) parameter specifies the number of iterations the AI model performs to refine and create the final output, with more steps resulting in more detailed and accurate outputs. The cfg (2.5) value, or "classifier-free guidance," controls how closely the generated output resembles the input prompt and reference information. This makes the virtual try-on result realistic while smoothly combining the chosen clothing piece onto the user's image. By taking advantage of these interrelated pieces and well-adjusted parameters, the system provides highly accurate, intuitive virtual try-on, facilitating online shopping by alleviating uncertainty and boosting customer confidence in their purchases.
- Send Image WebSocket returns the resultant processed image in PNG format to be displayed or reused. The image is transferred to the FastAPI server to maintain smooth data transmission and usability. The HTML pop-up gets the processed image from the server dynamically, presenting it in real-time to the user. This facilitates a smooth virtual try-on, where users get to view the clothing superimposed over their uploaded image.
- The image is then displayed in the pop-up, Were User gets an image with the selected clothing virtually applied to their photo.

III. System Requirements

Component	Minimum Requirement	Recommended Requirement
Cloud Service Provider	AWS / Google Cloud / Azure	AWS EC2 (GPU Instance)
Server Type	Virtual Machine (VM) / Dedicated Server	High-Performance GPU Server
Processor (CPU)	4-Core Intel Xeon / AMD EPYC	8-Core Intel Xeon Gold
Graphics Card (GPU)	NVIDIA T4 (16GB VRAM)	NVIDIA A100 (40GB VRAM)
RAM	16GB	32GB or higher
Storage (SSD)	100GB	1TB NVMe SSD
Operating System	Ubuntu 20.04 / Windows Server	Ubuntu 22.04 LTS
Networking	1 Gbps Bandwidth	10 Gbps for High-Speed Processing
Programming Language	Python 3.8+	Python 3.10+
Frameworks & Libraries	PyTorch, TensorFlow, FastAPI, OpenCV	Latest versions with CUDA acceleration
Database	PostgreSQL / MongoDB	Cloud-based PostgreSQL (AWS RDS)
Web Technologies	Node.js, React.js	Optimized Next.js / React.js
Security Measures	SSL Encryption, Cloud Firewalls	IAM Policies, GDPR Compliance

IV. PERFORMANCE ANALYSIS

A.Performance Analysis of CatVTON

To measure the performance of CatVTON, we compare its image accuracy, processing speed, and user satisfaction with the state-of-the-art virtual try-on models CP-VTON, TryOnGAN, and DeepVTON. These comparisons measure segmentation accuracy, realism, and system efficiency as a whole, indicating the advancements that the proposed model has made. The succeeding tables show a comprehensive performance analysis. [6]

Comparison of Image Accuracy Metrics:

Metric	CP-VTON	TryOnGAN	DeepVTON	CatVTON (Proposed)
IoU (Segmentation Accuracy)	80.1%	82.5%	84.3%	88.9%
SSIM (Realism Score)	0.89	0.90	0.91	0.93
Edge Consistency Score	82.4%	85.2%	86.8%	91.1%

Processing Speed Test Results:

Hardware Configuration	CP-VTON	TryOnGAN	DeepVTON	CatVTON (Proposed)
Local Machine (Intel i5, GTX 1660, 16GB RAM)	7.2 sec	6.8 sec	6.5 sec	5.1 sec
Cloud Server (NVIDIA A100, 32GB RAM)	3.2 sec	3.0 sec	2.9 sec	2.6 sec
AWS EC2 (T4 GPU, 16GB RAM)	4.8 sec	4.5 sec	4.2 sec	3.8 sec

Final Comparison Table:

Feature	CP-VTON	TryOnGAN	DeepVTON	CatVTON (Proposed)
Segmentation Accuracy (IoU)	80.1%	82.5%	84.3%	88.9%
Realism (SSIM Score)	0.89	0.90	0.91	0.93
Processing Speed (Avg. Time)	6.5 sec	5.8 sec	5.5 sec	2.9 sec
User Satisfaction	79%	82%	85%	90%

IV. BENEFITS

A.User engagement

Virtual Try-On improves customer experience through enabling users to virtually try products, thus making more informed buying decisions. The experience helps customers imagine how the products will fit and look prior to purchase, thus becoming more confident in what they choose. Further, studies have observed lower product return rates since consumers have a better idea of product fit and appearance before purchasing. This not only enhances customer satisfaction but also lowers return costs for retailers. Adopting Virtual Try-On technology can therefore result in a more interactive and personalized shopping experience, increasing customer loyalty and operational effectiveness [4].

The integration of Virtual Try-on technology will significantly improve the user engagement and purchasing decisions. By allowing customers to visualize products accurately, Virtual Try-on reduces uncertainty, leading to higher engagement and less return rates.

B.Effect on environment

In "The Intersection of Fashion, Immersive Technology, and Sustainability: A Literature Review," Patnaik et al. (2023) mention some of the challenges in the use of advanced technologies in the fashion industry. These are high energy consumption of digital infrastructures, consumer reluctance to indulge in new technologies, and lack of trained professionals to create and implement these systems. The research calls for an investigation to measure the real sustainability contribution of such technologies, track their adoption trends, and tackle the prevailing skills shortage in the sector.[5]

Custom Cart systems can be a major player in sustainability within the fashion world. Advanced AI technologies like Virtual Try-On play a key factor in lowering online returns, informing consumers, reducing waste in design and production, and lowering the need for physical products.

With Virtual Try-On functionality implemented, the Custom Cart platform will allow the users to see and visualize the products on themselves prior to buying, which results in more educated purchasing decisions and lower return rates. This not only improves the user's satisfaction but also reduces the environmental footprint linked with returns caused by transportation and overproduction of certain fashion products.

Additionally, by these virtual Try-on experiences, Custom Cart system minimizes the need for physical samples, thus reducing material waste and resource usage in the design process. Integrating this platform with such sustainable practices answers the environmental concerns and positions This system as an innovative solution within the fashion sector.

C. Competitive advantage

Virtual Try-On technology will provide companies with a competitive edge over others ecommerce spaces by providing an innovative ecommerce experience that does not exist on conventional e-commerce sites. In the midst of a very competitive e-commerce environment, where online retailers find it challenging to differentiate themselves, Virtual Try-On enables brands to differentiate by providing shoppers with an interactive and engaging way of discovering products. retailers can use this technology to draw a variety of shoppers from different age groups who are used to convenience and personalization from the reason ages, thus more visitors the store will have.

D. *Bridging the Gap between online and offline shopping*

One of the most significant challenges faced by the e-commerce industry is the impossibility to try on clothing products before the purchase to see how it looks on an individual hence Virtual Try-On addresses this problem. This solution along with the convenience of e-commerce space with the with Virtual Try-On feature will provide a more realistic and interactive shopping method, this feature will eliminate the feeling of procrastination of going to a store a picking a piece of clothing and increase purchasing satisfaction of a user, and makes online shopping as comfortable and exciting as going to a store.

V. LIMITATIONS

A. *High implementation cost*

Developing an original cutting-edge Virtual Try-On (VTON) system requires a combination of advanced AI algorithms, high-performance computing resources, and seamless e-commerce integration, all of which incur significant capital expenses. Software development is one of the principal cost factors as developing an AI-based try-on solution requires expertise in computer vision, machine learning, and web technologies, which translates to costly development and maintenance expenses. Apart from that, computation power and server hosting are needed in enabling real-time and seamless processing of virtual try-on pictures. Newer AI models like SAM (Segment Anything Model) and GroundingDINO require powerful GPUs and cloud AI capabilities, which small and medium retailers find challenging to accommodate.

Apart from computing power, data management and storage fees also form part of the overhead. Virtual try-on technologies must store and handle enormous amounts of user images, AI results, and model computations, which require safe, high-capacity cloud storage solutions. Moreover, some AI models and third-party APIs involve licensing fees, which are part of operational costs. All these reasons create it challenging for small businesses to implement AI-driven virtual try-on solutions at a huge cost. One such expense-cutting solution is to render AI models cost-effective, leverage edge computing to process locally, and utilize light-weight AI designs that minimize the requirement for high-cost cloud infrastructure.

B. *Lack of accuracy*

Even as AI-driven virtual try-on technologies aim to provide extremely realistic and accurate previews, they are not without a few accuracy-related issues that affect the user experience. One of the main constraints is the system's inability to adapt to varying light conditions. Under low-light or non-standard lighting conditions, AI models have been known to misrepresent colors of clothing, shadows, and textures, and thus produce unrealistic results. There is also another issue with shape and pose variability of the body because current segmentation models do not always properly accommodate clothing based on user shape, motion, or orientation, creating misaligned clothes or unnatural-appearing effects.

C. *Privacy concerns*

virtual try-on websites force users to upload their own photos, and this raises issues of data privacy and security threats. Storage of personal photos on cloud servers leaves them vulnerable to hacking, unauthorized use, and misuse. Users may also be left unaware of data storage period or if third parties have access to it, leading to potential privacy violation. In order to make sure that they are secure, sites ought to apply encryption, client-side processing, as well as explicit data retention policies to safeguard user data.

VI. FUTURE WORK

A. *Pre-uploaded Photo*

Adding an option for users to save a photo on their account can make virtual try-on easier for users and time saving. Instead of uploading a photo each time they try-on a product, users will be able to choose and upload a photo of their liking. This is especially helpful for more frequent shoppers allowing them to seamlessly experiment with different outfits without the process of uploading each time repetitive.

Additionally, the option to upload multiple photos in different poses and different lighting scenarios can enhance the versatility of the feature.

B. *Size input*

Being able to input body measurements or size (e.g., small, large or oversized) will enhance the accuracy of the generated photo to the actual product even more. This data can be used to adjust the fit of the actual clothing apparel to match real life proportions.

C. *Optimizing workflow*

The Comfy UI workflow can be further optimized to segment the clothing of the user faster and generate almost perfect masks by using larger segmentation models and using different nodes to provide faster results.

D. *Save option*

Giving the users the option to save the image on their devices will help the user to review the image later and compare it with other clothing. The user can also upload the photo to their social media accounts or send it to their friends, family members to ask about their opinion. This will enhance the user's satisfaction and confidence after the clothing product is received.

REFERENCES

[1] Wei Di, Anurag Bhardwaj, Vignesh Jagadeesh, Robinson Piramuthu and Elizabeth Churchill, “When relevance is not Enough: Promoting Visual Attractiveness for Fashion E-commerce” on Human-Computer Interaction (cs. HC), eBay Research Labs, June 2014.

[2] Zhuoyang Zhang, Han Cai and Song Han, “EfficientViT-SAM: Accelerated Segment Anything Model Without Accuracy Loss” onComputer Vision and Pattern Recognition (cs.CV); Artificial Intelligence (cs.AI); Machine Learning (cs.LG), 7 Feb 2024.

[3] Shilong Liu, Zhaoyang Zeng, Tianhe Ren, Feng Li, Hao Zhang, Jie Yang, Qing Jiang, Chunyuan Li, Jianwei Yang, Hang Su, Jun Zhu and Lei Zhang, "Grounding DINO: Marrying DINO with Grounded Pre-Training for Open-Set Object Detection" on Computer Vision and Pattern Recognition (cs.CV), Mar 2023.

[4] Akankshya Patnaik, Ashok Kumar Patnaik, Truptimayee Panigrahy, Maitrayee Rout, & Shiva Ram Patnaik. (2024). Exploring The Evolution Of Virtual Try-On Technologies: A Comprehensive Review

from a User-Centric Perspective. Educational Administration: Theory and Practice, 30(4), 8271–8287.

[5] Mesjar, Lyndsay, Karen Cross, Yang Jiang, and Josie Steed. 2023. "The Intersection of Fashion, Immersive Technology, and Sustainability: A Literature Review" *Sustainability* 15, no. 4: 3761.

[6] Zheng Chong, Xiao Dong, Haoxiang Li, Shiyue Zhang, Wenqing Zhang, Xujie Zhang, Hanqing Zhao, Dongmei Jiang & Xiaodan Liang , "CatVTON: Concatenation Is All You Need for Virtual Try-On with Diffusion Models" on Computer Vision and Pattern Recognition (cs.CV), Jul 2024.