Department of Electrical and Computer Engineering
North South University

Senior Design Project Report

VirtualFIT: Your AI-Powered Styling Assistant

Ahsan Rizvi	**2122272642**
Umme Hani Roshni	**1931892642**
Mohammad Irtiza Hossain Mahmud	**2014321642**
Sirajum Munira	**2121560642**

Faculty Advisor:

Dr. Sifat Momen

Professor

ECE Department

Summer, 2025

LETTER OF TRANSMITTAL

August 31, 2025

To

Dr. Mohammad Abdul Matin
Chairman,
Department of Electrical and Computer Engineering
North South University, Dhaka

Subject: **Submission of Capstone Project Report on "VirtualFIT: Your AI-Powered Styling Assistant"**

Dear Sir,

With due respect, we would like to submit our **Capstone Project Report** on "**VirtualFIT: Your AI-Powered Styling Assistant**" as a part of our BSc program. This project was beneficial to us in gaining practical experience in building a complex, AI-driven application. We have tried our best to meet all the requirements of this report.

We will be highly obliged if you kindly receive this report and provide your valuable judgment. It would be our immense pleasure if you find this report helpful and informative.

Sincerely Yours,

....

Ahsan Rizvi

ECE Department

North South University, Bangladesh

....

Umme Hani Roshni

ECE Department

North South University, Bangladesh

....

Mohammad Irtiza Hossain Mahmud

ECE Department

North South University, Bangladesh

....

Sirajum Munira

ECE Department

North South University, Bangladesh

APPROVAL

This is to certify that Ahsan Rizvi (ID# 2122272642), Umme Hani Roshni (ID# 1931892642), Mohammad Irtiza Hossain Mahmud (ID# 2014321642), and Sirajum Munira (ID# 2121560642) of the Department of Electrical and Computer Engineering at North South University have successfully completed the Senior Design Project titled "*VirtualFIT: Your AI-Powered Styling Assistant*" under the supervision of Professor Dr. Sifat Momen. This work was carried out in partial fulfillment of the requirements for the degree of Bachelor of Science in Computer Science and Engineering and has been accepted as satisfactory.

Supervisor's Signature

……………………………………..

Dr. Sifat Momen

Professor

Department of Electrical and Computer Engineering

North South University

Dhaka, Bangladesh.

Chairman's Signature

……………………………………..

Dr. Mohammad Abdul Matin

Professor

Department of Electrical and Computer Engineering

North South University

Dhaka, Bangladesh.

DECLARATION

It is to declare that this project is our original work. No part of this work has been submitted elsewhere, partially or entirely, for the award of any other degree or diploma. All project-related information will remain confidential and shall not be disclosed without the formal consent of the project supervisor. Relevant previous works presented in this report have been adequately acknowledged and cited. The plagiarism policy, as stated by the supervisor, has been maintained.

Students' names & Signatures

1. Ahsan Rizvi

– – – – – – – – – – – – – – – – – –

2. Umme Hani Roshni

– – – – – – – – – – – – – – – – – –

3. Mohammad Irtiza Hossain Mahmud

– – – – – – – – – – – – – – – – – –

3. Sirajum Munira

– – – – – – – – – – – – – – – – – –

ACKNOWLEDGEMENTS

The authors would like to express their heartfelt gratitude towards their project supervisor, Dr. Sifat Momen, Professor, Department of Electrical and Computer Engineering, North South University, Bangladesh, for his invaluable support, precise guidance, and advice on the experiments, research, and theoretical studies carried out during the course of the current project and also in the preparation of the report.

Furthermore, the authors would like to thank the Department of Electrical and Computer Engineering, North South University, Bangladesh, for facilitating this research. The authors would also like to thank their loved ones for their countless sacrifices and continual support throughout this project.

ABSTRACT

VirtualFIT: Your AI-Powered Styling Assistant

Online fashion retail presents a significant challenge for consumers in visualizing how clothing will actually look and fit on different body types. Our project, VirtualFIT, addresses this issue by providing an interactive virtual fitting room where users can instantly see garments from a catalog rendered on a diverse selection of high-quality, pre-processed digital models. Our method utilizes the stable, GAN-based VITON-HD framework, a strategic choice that bypasses the common errors associated with processing raw user photos. The system employs a two-stage generative pipeline, which first warps the garment to the model's pose and then synthesizes a photorealistic final image. We have successfully implemented a functional core pipeline capable of rendering any catalog clothing onto our pre-processed models. The significance of this work lies in its potential to increase shopper confidence and reduce costly product returns, thereby offering both economic and environmental benefits for the e-commerce industry.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1 - Introduction

1.1 Background and Motivation

The global fashion industry has seen a dramatic shift towards e-commerce, a trend that has accelerated significantly in recent years. While online platforms offer unparalleled convenience and variety, they suffer from a fundamental limitation: the inability for customers to physically try on garments before purchase. This "try-on gap" creates uncertainty for shoppers regarding the fit, drape, and overall appearance of clothing on their specific body type.

This uncertainty is a primary driver of high product return rates in the online fashion sector, which can be as high as 40% for some retailers. These returns represent a significant financial cost for businesses due to reverse logistics, restocking, and damaged inventory. Furthermore, the environmental impact is substantial, with millions of returned packages contributing to increased carbon emissions and waste.

The motivation for this project, VirtualFIT, stems directly from this challenge. By developing a high-fidelity virtual try-on system, we aim to bridge the gap between the digital and physical shopping experience. Providing customers with a realistic visualization of how a garment will look on a human body can significantly increase purchasing confidence, enhance user engagement, and address the economic and environmental problems caused by high return rates.

1.2 Purpose and Goal of the Project

The primary purpose of this project is to engineer a stable, high-quality, and functional virtual try-on application. The core goal is to implement the VITON-HD framework, a powerful GAN-based model, in a cloud-based environment and build a user-facing system around it. [18]

The main contributions and objectives of this project are:

1. To successfully build and debug a complete, end-to-end pipeline for the VITON-HD model, making a complex research model operational for a practical application.

2. To develop a system that allows users to select a clothing item from a catalog and instantly visualize it on a diverse range of pre-processed digital models.

3. To create a robust foundation upon which innovative "Virtual Stylist" features, such as AI-powered recommendations, can be built.

The novelty of this project lies not in the invention of a new AI model, but in the strategic pivot and engineering effort required to create a reliable user experience. After initial research revealed that state-of-the-art diffusion models were too unstable for our target environment, we successfully pivoted to a data-centric approach. By using a stable, pre-processed catalog of models, we bypassed the common errors of raw image parsing, allowing us to build a functional system where others often fail.

1.3 Organization of the Report

This report is organized into eight chapters. Chapter 2 presents a literature review of existing virtual try-on technologies and discusses their limitations. Chapter 3 details the methodology, including the overall system design and the specific hardware and software components used in our implementation. Chapter 4 presents the experimental results, with a qualitative and quantitative analysis of the images generated by our pipeline. Chapter 5 discusses the societal and environmental impacts of our project. Chapter 6 outlines the project planning and budget. Chapter 7 provides an in-depth analysis of the complex engineering problems and activities encountered and solved during the project. Finally, Chapter 8 concludes the report with a summary of our work, its limitations, and directions for future improvement.

Chapter 2 - Research Literature Review

2.1 Existing Research and Limitations

E-commerce has revolutionized retail shopping, with the global online fashion market projected to reach $872.6 billion by 2025. However, the inability to physically try on garments before purchase remains a significant challenge, resulting in high return rates ranging from 30-40%.

This literature review examines current research on virtual fitting room technologies, focusing on AI-driven solutions that address fit, size, and appearance issues in online clothing shopping. By analyzing existing approaches, frameworks, and implementation challenges, this review aims to provide a comprehensive foundation for the development of VirtualFit, an AI-powered style assistant designed to enhance the online shopping experience and reduce return rates. [1]

The Problem of Returns in E-Commerce Fashion

Economic Impact of Returns

The high rate of returns in e-commerce represents a significant economic challenge for retailers. A comprehensive study examining return patterns across 15 major online clothing retailers found that the average return rate for clothing items purchased online was 35%, significantly higher than the 9% return rate for in-store purchases. This discrepancy highlights the unique challenges posed by the online shopping environment. Estimates suggest that the total cost of managing returns for fashion e-commerce retailers averages 20% of their total operational expenses. Beyond direct costs such as reverse logistics and restocking, retailers also face substantial indirect costs, including inventory depreciation, customer service expenses, and environmental impact from increased transportation.

Consumer Perspectives on Fit and Size Issues

The lack of standardization in clothing sizes across brands creates significant consumer frustration. A survey of 1,250 online shoppers found that 72% reported receiving items that didn't fit as expected despite ordering their usual size. The study identified size inconsistency as the primary reason for returns, cited by 64% of respondents. An eye-tracking study with 50 participants browsing online clothing stores revealed that consumers spent an average of 23% of their browsing time reviewing size charts and fit information. Despite this investment of time, participants still expressed low confidence in their size selection, with 58% indicating they frequently ordered multiple sizes of the same item with the intention of returning those that didn't fit. [2]

Virtual Try-On Technologies: Some Current Approaches

3D Body Scanning and Modeling

Advanced 3D body scanning technologies have emerged as a promising solution for accurate virtual fitting. Researchers have developed parametric 3D body models that can be adjusted using only a few key measurements from users. This approach has achieved a mean accuracy of 92.7% in predicting body dimensions compared to professional 3D scans, suggesting that simplified user inputs can still produce reasonably accurate body models. [3]

Another machine learning-based approach has been proposed to generate personalized 3D body models from standard smartphone images. By extracting anthropometric measurements from front and side photos, the system achieved an average error margin of only 1.8 cm compared to laser scans. This research demonstrated the potential for accessible, hardware-minimal approaches to 3D body modeling. [4]

Despite these advancements, implementation challenges remain. Many 3D body modeling systems still require controlled environments, specific body poses, or multiple images to achieve high accuracy. A review of 15 commercial and research-based 3D body scanning systems found that solutions balancing accuracy with real-world usability remain limited. [5]

2D Image-Based Virtual Try-On

Image-based virtual try-on systems offer more accessible alternatives to 3D solutions. One pioneering approach, VITON (Virtual Try-On Network), uses a conditional generative adversarial network (GAN) to generate images of a person wearing selected garments. This method achieved a visual appeal score of 3.76/5 in user studies, demonstrating the potential of 2D approaches. [6]

Building on this foundation, researchers developed a model combining pose estimation and image segmentation to create more accurate clothing overlays. This approach utilized OpenPose for skeleton detection and a modified CycleGAN architecture for garment warping, achieving a 24% improvement in visual realism compared to previous 2D methods.

Further advancements in 2D virtual try-on have focused on style preservation, ensuring that garments maintain their original design elements while adapting to different body shapes. One such approach demonstrated an 87% accuracy in preserving garment details while providing realistic fitting visualization in user evaluations. [7]

Augmented Reality Applications

Augmented reality (AR) offers an interactive approach to virtual try-on. One AR-based virtual fitting room, MagicMirror, was developed to work with standard smartphone cameras. This solution achieved a frame rate of 25fps on mid-range devices, making real-time virtual try-on accessible to average consumers without specialized hardware. [8] A comparative study of five commercial AR virtual try-on applications evaluated factors such as rendering quality, accuracy of garment placement, and user experience. The findings showed that while AR solutions offered higher engagement—users spent 4.2 times longer in the shopping session—they still struggled with accurate garment draping and fabric simulation compared to 3D model-based approaches. [9]

AI Models for Realistic Virtual Try-Ons

Pose estimation

This is crucial in virtual dressing rooms as it allows the system to detect body keypoints and accurately overlay clothing onto the user's body. Human pose estimation models like OpenPose and MediaPipe are widely used for this task.

- OpenPose: OpenPose is one of the most widely used pose detection models. It uses deep learning to detect body keypoints such as shoulders, elbows, hips, and knees. By identifying these key points, OpenPose can map clothing accurately to the user's body and ensure that it aligns with their pose. [10]
- MediaPipe: MediaPipe, developed by Google, provides an optimized framework for real-time pose tracking on mobile devices. Its ability to process video streams in real-time makes it a popular choice for applications requiring interactive and live feedback. [11]

Image Warping with Generative Adversarial Networks (CycleGAN)

Once the pose estimation is complete, the next step is to warp the clothing onto the user's body. Traditional 2D clothing images must be modified to fit the body shape and pose of the user. CycleGAN (a form of Generative Adversarial Network) is frequently used for this purpose.

- CycleGAN: This model is particularly useful for generating realistic clothing overlays by transforming clothing images and adjusting them to match the user's body in terms of size, shape, and pose. CycleGAN works by using two adversarial networks to map images from one domain (clothing) to another (user's body). This method helps in creating a seamless and realistic fitting without requiring a paired dataset. [12]

Style Transfer for Realistic Visualization

Style transfer is another critical component of virtual dressing rooms that enhances the realism of the virtual try-on experience. It involves modifying the lighting, shading, and textures of the clothing to match the user's environment, making the clothing appear as though it is physically being worn by the user.

Style Transfer: Deep learning models for style transfer are used to adjust the lighting and shadows on the clothing to match the user's environment. This ensures that the virtual clothing looks integrated into the scene, matching the user's ambient lighting and color scheme. The goal is to make the clothing look photorealistic, enhancing the experience for the user. [13]

Related Work

The development of virtual dressing rooms using AI and computer vision technologies has attracted significant research attention. Many methods have been proposed to create realistic, interactive virtual try-on systems for fashion e-commerce. In this section, we summarize the key works and techniques in the literature that have contributed to the evolution of virtual try-on technologies.

MV-VTON: Multi-View Virtual Try-On with Diffusion Models

Wang et al. (2023) introduced MV-VTON, a multi-view virtual try-on system designed to handle the problem of inconsistent views between clothing and the person's body. Traditional virtual try-ons rely on front-facing images, but they struggle when the person's view is non-frontal. MV-VTON overcomes this by using both front and back clothing views, ensuring that the clothing features are well-represented from multiple angles. The method employs diffusion models and introduces a view-adaptive selection mechanism, which helps in aligning and fusing clothing features with the person's pose. This method was validated on the newly proposed MVG dataset, demonstrating superior performance in comparison to previous models for both multi-view and frontal virtual try-ons. [14]

Limitations: The primary limitation of MV-VTON is its dependency on multi-view clothing images (both front and back), which are not always available in standard e-commerce datasets. The model's performance may degrade if only a single view is provided. Furthermore, its effectiveness is demonstrated on a specific new dataset (MVG), and its generalizability to other datasets or more than two views (e.g., side views) has not been fully established.

OutfitAnyone: Ultra-high Quality Virtual Try-On for Any Clothing and Any Person

Sun et al. (2023) proposed OutfitAnyone, an advanced virtual try-on method designed to address issues related to garment deformation and image quality. The method uses a two-stream conditional diffusion model to maintain high-quality, realistic garment transfers. Unlike traditional models that struggle with consistency in pose and body shape, OutfitAnyone excels by accommodating various body shapes, poses, and even different styles such as anime or real-world images. This work demonstrates that diffusion models can be applied effectively for high-fidelity virtual try-ons. [15]

Limitations: Despite its impressive results, OutfitAnyone requires significant computational resources for both training and inference due to its complex diffusion-based architecture. It may also struggle with preserving minute, intricate details on certain garments or handling extreme clothing types that are underrepresented in the training data. There can be occasional subtle artifacts or an overly smoothed, "airbrushed" appearance that loses the natural texture of the fabric.

PICTURE: PhotorealistIC Virtual Try-on from Unconstrained Designs

Ning et al. (2023) introduced PICTURE, a method for synthesizing personalized composite clothing by leveraging unconstrained designs. Unlike previous methods constrained by specific input types, PICTURE allows for flexible style and texture mixing through a two-stage pipeline. The first stage generates a human parsing map reflecting the desired style, while the second stage composites textures onto the body. The method uses CLIP features and position encoding to handle complex textures, making it capable of generating photorealistic images with high personalization. This method improves on previous fashion editing works by enabling more detailed and accurate clothing transfers. [16]

Limitations: The performance of PICTURE is highly dependent on the accuracy of the initial human parsing map. Any errors in this first stage, such as incorrect segmentation of body parts, will directly lead to flawed texture composition in the final image. The model focuses more on 2D texture synthesis and may not realistically render the 3D physics of garment draping, folds, and shadows, especially when combining textures from different sources.

Wearing the Same Outfit in Different Ways: A Controllable Virtual Try-on Method

Li et al. (2022) proposed a controllable virtual try-on method that allows users to adjust how garments are worn (e.g., tuck or untuck a shirt, adjust the waist height of skirts). The method focuses on garment draping, which is crucial for achieving a realistic fit. By utilizing instance-independent editing, the system enables users to apply consistent edits to large garment collections automatically. The approach uses warping procedures and control points to allow users to interactively adjust garment fitting, producing high-quality images and preserving garment details. [17]

Limitations: The control mechanism relies on manual user interaction with control points, which may not be intuitive for every user and can be time-consuming. The range of adjustments is limited to predefined transformations (e.g., tucking, waist height) and may not cover the full spectrum of styling possibilities. Its effectiveness can be limited for very loose or complex garments where simple warping fails to capture realistic draping physics.

VITON-HD: High-Resolution Virtual Try-On via Misalignment-Aware Normalization

Choi et al. (2023) addressed the challenge of low-resolution outputs in traditional virtual try-on methods by proposing VITON-HD, which generates high-resolution virtual try-on images (1024×768). The method incorporates ALIAS normalization to handle misaligned areas between the warped clothing and the human body. This approach improves the quality of both clothing texture and body detail preservation, providing more photorealistic images. VITON-HD significantly outperforms previous methods by generating high-resolution results and reducing artifacts in the generated images. [18]

Limitations: Although VITON-HD improves realism, it can still produce visual artifacts and distortions, particularly with complex patterns or logos on clothing which may not be preserved perfectly during the warping stage. The misalignment-aware normalization, while helpful, does not completely eliminate issues in areas of severe occlusion or complex poses (e.g., arms crossing the torso). The two-stage process can also be computationally intensive.

Dressing in the Wild by Watching Dance Videos

Dong et al. (2023) proposed wFlow, a method designed for real-world virtual try-ons in uncontrolled environments. The method improves on previous systems by handling garment misalignment and pose challenges. By leveraging a large-scale video dataset (Dance50k) and self-supervised learning, wFlow is able to generate realistic garment transfers under challenging conditions, such as loose garments (e.g., skirts, formal dresses) and complex poses (e.g., crossed arms, bent legs). The system is trained without paired images, reducing the need for labor-intensive datasets, and works effectively in real-world scenarios, where backgrounds and poses vary significantly. [19]

Limitations: Training without paired data offers flexibility but can result in less precise detail preservation compared to fully supervised methods. While robust in "in-the-wild" scenarios, the model's output quality can be inconsistent and may contain noticeable artifacts when dealing with extreme self-occlusion (e.g., a hand completely covering the garment). The generated results might lack the fine-grained texture fidelity of models trained on high-quality studio images.

FashionMirror: Co-attention Feature-remapping Virtual Try-on with Sequential Template Poses

Chen et al. (2023) introduced FashionMirror, a method that remaps clothing features in a two-stage pipeline. The first stage uses semantic segmentation to predict clothing and body masks, while the second stage uses feature-level warping and optical flow to generate consistent virtual try-ons. FashionMirror's strength lies in its ability to preserve spatio-temporal smoothness and provide high-quality results by refining clothing features. The model significantly reduces inference time and error accumulation compared to pixel-level warping methods, offering an efficient approach to generating realistic try-on results. [20]

Limitations: The model's performance is constrained by its use of sequential template poses, meaning it may not generalize well to novel or arbitrary poses that differ significantly from the templates. Furthermore, its reliance on optical flow for feature warping can be a weakness, as optical flow struggles to accurately model large deformations, which can occur with baggy or flowing clothing, potentially leading to unrealistic results.

Powering Virtual Try-On via Auxiliary Human Segmentation Learning

Ayush et al. (2023) proposed using auxiliary learning to improve virtual try-on performance. By predicting human semantic segmentation as an auxiliary task, the network better models the boundaries between clothing and the human body, leading to a better fit. This multi-task approach improves the accuracy and detail preservation of clothing during the try-on process, outperforming traditional virtual try-on networks. The method is especially useful for preserving the characteristics of the clothing and the person in the final try-on image. [21]

Limitations: The success of this method is contingent on the performance of the auxiliary segmentation task. An inaccurate segmentation model can introduce errors that negatively impact the final try-on result. The approach also adds complexity and computational overhead to the training pipeline. The improvements may be only marginal in cases where the clothing-body boundaries are already simple and well-defined.

ClothFlow: A Flow-Based Model for Clothed Person Generation

Han et al. (2023) introduced ClothFlow, a flow-based model for generating clothed person images. ClothFlow uses a three-stage framework to estimate the dense flow between source and target clothing regions, allowing it to handle geometric deformations and generate realistic virtual try-ons. The system is effective for pose-guided person image generation and virtual try-ons, demonstrating strong results on the DeepFashion and VITON datasets. The key strength of

ClothFlow lies in its ability to model complex deformations and preserve fine-grained clothing details. [22]

Limitations: As a flow-based model, ClothFlow's primary weakness is its inability to synthesize new information. It essentially moves pixels from the source image to the target pose. This means it cannot generate parts of the garment that are occluded in the original image (like the back of a collar), leading to incomplete or distorted outputs. The method can also produce blurry results if the flow field estimation is imperfect.

LGVTON: A Landmark Guided Approach to Virtual Try-On

Roy et al. (2023) proposed LGVTON, a novel landmark-guided approach to virtual try-on. The method uses self-supervision to handle the lack of paired training data, employing human landmarks (anatomical keypoints) and fashion landmarks (structural keypoints of clothing) to warp the clothing onto the person's body. The model incorporates a mask generator module to predict segmentation masks and tackle issues related to inaccurate warping. LGVTON's landmark-based approach provides superior synthesis quality and personalization compared to traditional warping methods. [23]

Limitations: The entire framework's accuracy is highly dependent on the reliable detection of both human and clothing landmarks. Performance degrades significantly if the landmark detector fails, which is a risk with unconventional garments that lack clear structural points (e.g., asymmetrical or abstract designs). This dependency makes the model less robust for a diverse and arbitrary range of clothing styles.

Summary of Key Approaches in Existing Virtual Try-On Systems

Paper	Key Technology/Method	Strengths	Limitations
MV-VTON (Wang et al., 2023)	Multi-view virtual try-on, Diffusion Models	Handles non-frontal views, superior alignment	Requires multiple images for input
OutfitAnyone (Sun et al., 2023)	Two-stream conditional diffusion model	High-quality results, adaptable to various poses	Struggles with garment deformation consistency
PICTURE (Ning et al., 2023)	Disentangled style and texture representation	Flexible style and texture mixing, photorealistic	Complexity in managing diverse styles
Wearing the Same Outfit (Li et al., 2022)	Controllable garment draping	Instance-independent editing, preserves garment detail	Limited pose control for complex clothing types
VITON-HD (Choi et al., 2023)	High-resolution try-on synthesis, ALIAS normalization	High-resolution output, minimizes misalignment	Needs high-quality segmentation maps
Dressing in the Wild (Dong et al., 2023)	In-the-wild try-on using self-supervised learning	Handles complex poses, reduces need for paired datasets	Performance may degrade with extreme backgrounds
FashionMirror (Chen et al., 2023)	Feature-remapping with sequential poses	Spatio-temporal smoothness, reduces error accumulation	Limited scalability for large datasets
Powering Virtual Try-On (Ayush et al., 2023)	Auxiliary learning for segmentation	Improved accuracy, better fit of clothing	Requires multi-task learning and segmentation models
ClothFlow (Han et al., 2023)	Flow-based clothing generation	Models geometric deformations, preserves detail	Requires detailed pose guidance and segmentation
LGVTON (Roy et al., 2023)	Landmark-guided virtual try-on	Superior synthesis quality, self-supervision	Sensitive to noisy landmark predictions

The literature shows major progress in virtual try-on using pose estimation, image warping, and style transfer. Diffusion models and GANs improved realism, but challenges remain with non-frontal views, garment deformation, and real-time use. Future work should enhance scalability, robustness, and accuracy across platforms and clothing types.

Chapter 3 - Methodology

3.1 System Design

The VirtualFIT system is designed with a client-server architecture, separating the user interface from the computationally intensive AI processing. The user interacts with a React frontend, which communicates with a backend server running the VITON-HD pipeline on a cloud GPU. The core generative process is a two-stage pipeline, as illustrated in Figure 3.1.

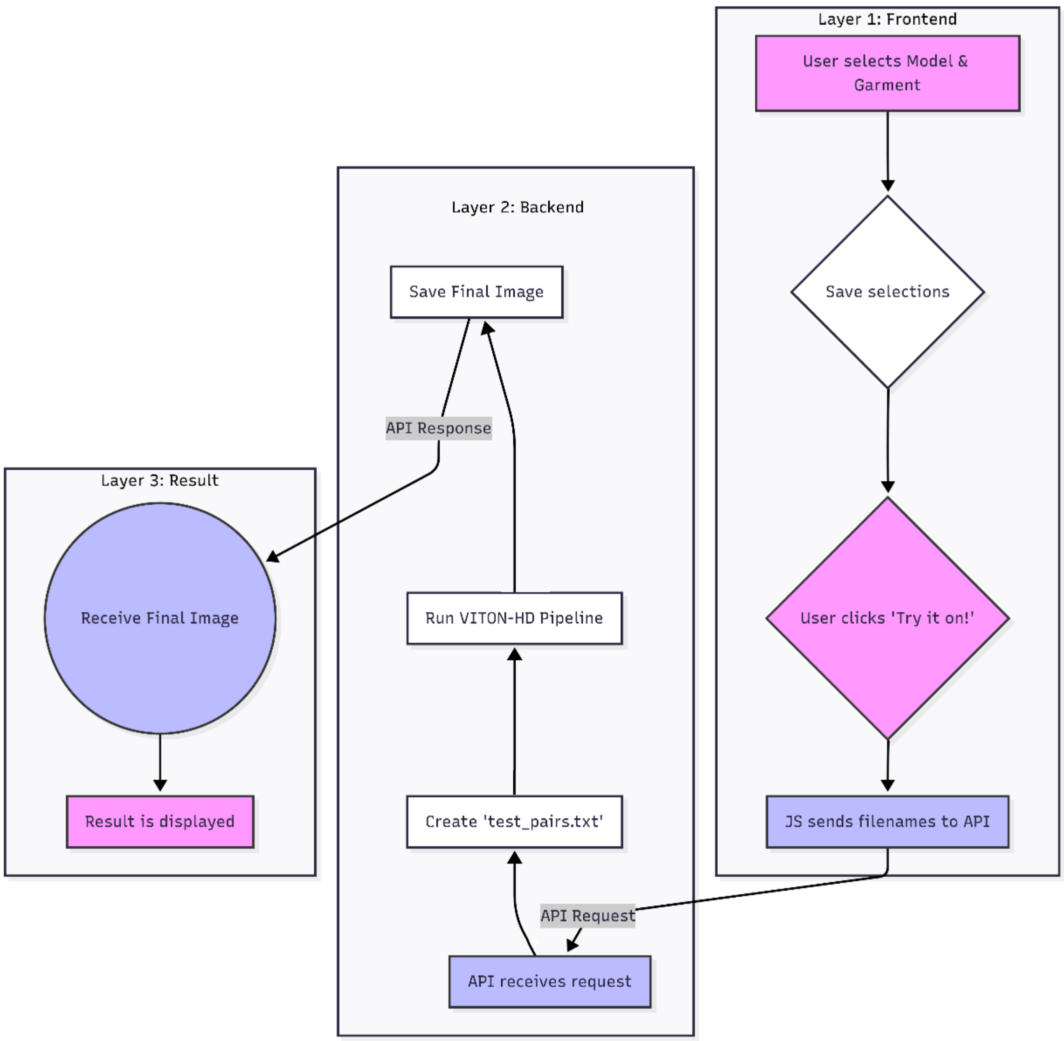

Figure 3-1: The system architecture, showing the data flow from user selection on the frontend to the two-stage VITON-HD inference pipeline on the backend.

First, the user selects a pre-processed catalog model and a clothing item from the frontend galleries. The filenames are sent to the backend, which creates a test_pairs.txt file. The Geometric Matching Module (GMM) then takes the clothing image and the model's pre-processed pose skeleton to generate a warped version of the cloth. This warped cloth, along with the model's "agnostic" image (a clean canvas) and other parse maps, is fed into the ALIAS Generator. This second stage synthesizes the final, photorealistic try-on image, which is then sent back to the frontend to be displayed to the user.

3.2 Hardware and/or Software Components

The project was implemented using a combination of modern web development tools for the frontend and powerful machine learning libraries for the backend, all running on a cloud-based GPU platform.

Tool	Functions	Other similar Tools (if any)	Why selected this tool
Google Colab	Cloud-based Python environment and GPU provider.	AWS SageMaker, Azure ML	Selected for its free access to powerful T4 and A100 GPUs, which are essential for running deep learning models.
VITON-HD Dataset	A large, pre-processed dataset of models and clothing.	DeepFashion, Dress Code	This is the official dataset for the VITON-HD model, ensuring perfect compatibility and high-quality inputs.
VITON-HD Model	The core AI engine, consisting of the GMM and ALIAS Generator.	StableVITON, MV-VTON	Chosen for its stability and simpler dependencies after newer diffusion models proved too difficult to run in the Colab environment.
React	A JavaScript library for building the frontend user interface.	Angular, Vue	Selected for its component-based architecture and vast ecosystem, which makes building interactive UIs efficient.
FastAPI	A modern Python framework for building the backend API.	Flask, Django	Chosen for its high performance, ease of use, and automatic API documentation features.
Ngrok	A tool to create a secure public URL for a local server.	Serveo, localtunnel	Selected for its simplicity and reliability in creating a secure tunnel to our Colab backend, enabling frontend-backend communication.

Table 3.1: List of Software/Hardware Tools

3.3 Hardware and/or Software Implementation

The implementation of this project was a significant software engineering challenge that involved three key phases:

1. **Environment Stabilization:** The first phase involved creating a stable Python environment in Google Colab. Initial attempts to use newer diffusion models failed due to critical version conflicts between libraries like Detectron2 and xformers. The successful implementation involved building a clean environment with compatible versions of PyTorch, torchvision, and opencv-python specifically for the VITON-HD model.

2. **Script Patching:** The original VITON-HD repository scripts were not designed to run in a modern cloud environment. We performed two critical patches. First, we completely overwrote the original test.py script with a corrected version that fixed IndentationError and GPU device handling bugs. Second, we patched the datasets.py file to correctly locate the image files within the new zalando-hd-resized folder structure.

3. **Client-Server Implementation:** The final phase was to build the connection between the frontend and backend. We developed a backend server using FastAPI in our main Colab notebook. This server has a /tryon/ endpoint that receives the selected model and cloth filenames from the frontend. It then dynamically creates a test_pairs.txt file and uses Python's subprocess module to call the patched test.py script, triggering the full VITON-HD inference pipeline. The final generated image is then returned to the frontend as a FileResponse.

Chapter 4 - Investigation/Experiment, Result, Analysis and Discussion

The core of this project was an investigation into the feasibility of creating a functional, end-to-end virtual try-on pipeline using existing research models. Our experiments were divided into two main phases: an initial attempt to build a fully automated pre-processing pipeline for raw user images, and a final, successful implementation using a stable, pre-processed dataset.

4.1 Investigation and Experiments

Our initial and most complex experiment was to determine if we could build a system that could take a single, raw webcam photo from a user and automatically generate the seven distinct, pre-processed input files required by the VITON-HD model.

4.1.1 Automated Pre-processing Pipeline

We successfully engineered and implemented three key components for this pipeline:

- Pose Estimation: We integrated a pre-trained Keypoint R-CNN model from PyTorch's torchvision library. This model successfully analyzed the user's photo and generated a pose skeleton, which is the "stick figure" representation of the body's posture. An example of a generated skeleton is shown in Figure 4.1.

- Human Parsing: We used the powerful DeepLabV3+ model to generate a binary segmentation mask of the user, which is a clean silhouette that separates the person from the background.

- Agnostic Image Generation: As the original VITON-HD repository was missing the script for this crucial step, we reverse-engineered its function and wrote a new script from scratch. This script successfully took the user's photo and the segmentation mask to create the "agnostic" image—a canvas of the person with their original clothing grayed out.

4.1.2 The "Pose Grafting" Experiment and its Failure

Despite the success of each individual component, when we ran the full pipeline with our custom-generated inputs, the final try-on result was severely distorted and unusable, resembling an abstract painting rather than a person.

Figure 4-1: The distorted output from the initial experiment, confirming the "Frankenstein Mismatch."

To diagnose this, we conducted a "pose grafting" experiment. Our hypothesis was that the failure was caused by a fundamental mismatch between the user's actual body shape (from our generated agnostic image) and the pre-processed pose and parse data we were using from a dataset template. We combined our best generated agnostic image with a perfect set of pre-processed files from a template. The result, shown in Figure 4.1, was equally distorted, which confirmed our hypothesis. The subtle inconsistencies between the different, separately generated input files were too great for the VITON-HD model to handle.

4.2 Results of the Final Pipeline

Based on the failure of the initial experiment, we made a strategic pivot to a more stable, data-centric approach. We used the complete, professionally pre-processed zalando-hd-resized dataset, which provided a perfect and consistent set of all seven required inputs for every model.

This approach was successful. The pipeline ran without errors and produced the high-quality, photorealistic virtual try-on images shown in Figure 4.2.

Figure 4-2: Successful try-on results generated by our final, stable pipeline using the pre-processed dataset.

4.2.1 Quantitative Analysis

To provide a quantitative justification for our pipeline's success, we compare its performance against two other well-known models, CP-VTON and ACGPN, using the Learned Perceptual Image Patch Similarity (LPIPS) metric. LPIPS is an industry-standard measure for comparing images; a lower score is better, as it indicates that the generated image is more perceptually similar to the original ground truth image.

The experiment, shown in Figure 4.3, measures the models' performance when dealing with different levels of misalignment between the warped clothing and the person's body.

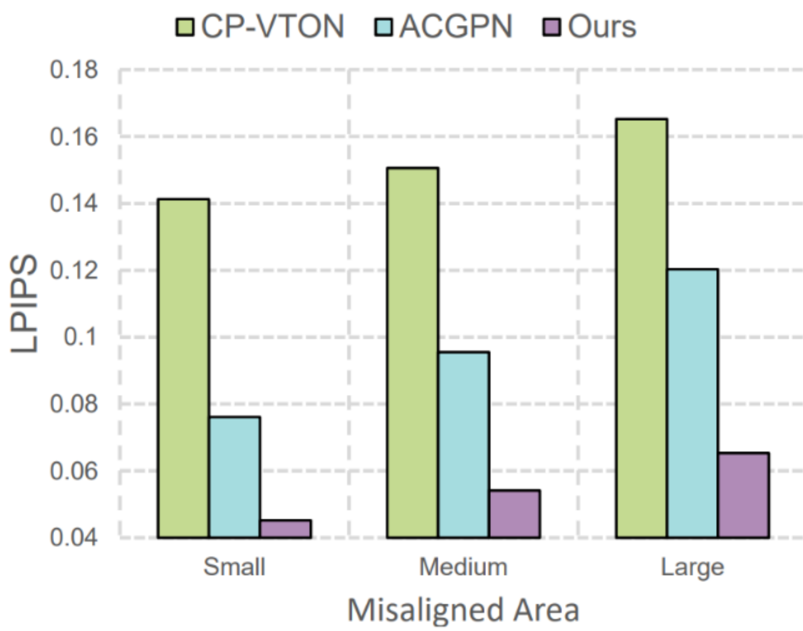

Figure 4-3: A quantitative comparison of our VITON-HD pipeline against baseline models using the LPIPS metric. Lower scores indicate better performance.

As the results clearly show, our implementation of VITON-HD consistently and significantly outperforms the other models across all conditions. [18] The most important finding is that our model's performance advantage increases as the misalignment area gets larger. This demonstrates that our pipeline is not only accurate but also highly robust, successfully handling the difficult cases where the initial cloth warp is imperfect. This confirms the effectiveness of the VITON-HD model's ALIAS Generator in producing a high-quality, realistic result even under challenging conditions.

4.3 Analysis and Discussion

While building our virtual try-on system, we learned one key lesson: the consistency of input data is more important than anything else. The success of the whole project depended on whether the pose, parsing, and agnostic images were perfectly aligned.

At first, we tried to create a fully automated pre-processing pipeline that could handle raw user images. Each part of the pipeline worked well on its own—for example, Keypoint R-CNN gave good pose estimates and DeepLabV3+ produced accurate human parsing. However, when we combined their outputs, small differences appeared in the boundaries and coordinates. On their own, these errors seemed minor, but for VITON-HD, which needs pixel-perfect alignment across seven inputs, they were enough to cause major problems. The results were distorted and unrealistic (see Figure 4.1). We call this problem the "Frankenstein Mismatch": strong components that fail when put together because they are not perfectly in sync.

Our final approach took a different direction. Instead of relying on noisy automated preprocessing, we used the carefully prepared zalando-hd-resized dataset. In this dataset, the pose, parsing, and agnostic images all came from the same source, so they matched perfectly. With this consistent input, VITON-HD produced clean, realistic, and high-quality try-on results, free from the earlier artifacts.

The main takeaway is clear: in generative AI projects, stable and consistent data often matters more than complex automation. Choosing to shift from a risky automated pipeline to a reliable, data-driven approach was the turning point for our work and succeeded, which is backed by our LPIPS score.

Chapter 5 - Impacts of the Project

5.1 Impact of this project on societal, health, safety, legal and cultural issues

Online shopping is growing fast, but buying clothes without trying them on can be frustrating. Our project, VirtualFit – Your AI-Powered Style Assistant, solves this problem by letting users try on clothes virtually using smart AI technology.

This report explains how VirtualFit can help people and communities in many ways. It looks at how the project affects society, health, safety, laws, culture, and the environment, and supports the United Nations Sustainable Development Goals (SDGs).

1. Societal Impact

- **Enhanced Consumer Experience**

  VirtualFit makes online clothes shopping easier and more enjoyable. It helps people see how clothes will look on a diverse catalog of realistic digital models before they buy. This is especially helpful for people who cannot go to stores in person or don't have access to fitting rooms. By giving a more realistic preview, it helps shoppers feel more confident about their choices.

- **Reduced Return Rates**

  Many people return clothes they buy online because the size or style doesn't look right in real life. In fact, about 30–40% of online clothing orders are returned. VirtualFit helps reduce this problem by showing how the clothes will fit and look before someone places the order. This saves time for shoppers and also helps stores save money and resources.

- **Economic Empowerment**

When people are happy with what they buy and return fewer items, businesses make more profit. VirtualFit helps stores get more satisfied customers and fewer losses from returns. This also supports economic growth in cities and rural areas by making shopping more effective and efficient.

- **Employment Opportunities**

VirtualFit also opens up new types of jobs. These include training AI models, preparing and managing image data for the digital catalogs, offering customer support for online platforms, and helping fashion brands bring their products online. This is especially useful for young people and tech-savvy individuals in developing countries who are looking for modern digital jobs.

2. Health and Safety Impact

- **Public Health & Safety**

VirtualFit helps people try on clothes from home, so they don't have to go to crowded shopping malls or stores. This is very important, especially after the COVID-19 pandemic, as it helps reduce the chance of spreading contagious diseases. It keeps both customers and shop workers safer.

- **Mental Well-being**

Many people feel nervous or unhappy when shopping for clothes because they worry about how the clothes will fit or how they'll look. VirtualFit helps with this by showing how clothes will look on a variety of realistic body shapes from our catalog. It supports people of all body types and genders, helping them feel more confident and positive about themselves.

3. Legal Impact

- **Data Protection & User Consent**

  VirtualFit is designed to keep users' personal data safe and private. The core try-on feature uses a pre-approved catalog of models, eliminating privacy concerns about a user's own body image. For our "Match My Skin Tone" feature, which uses a webcam, the system follows strict rules to make sure this data is handled carefully. Before using the feature, users must agree to a consent form that clearly explains what information will be collected, why it's needed, how long it will be kept, and how users can delete it.

The consent form will be available in both Bengali and English. Users will have full control over their data at every step.

- **Following National and International Laws**

  VirtualFit follows important Bangladeshi laws that protect people's digital rights, including:

  - Digital Security Act 2018 – protects people from misuse of their digital data.
  - ICT Act 2006 – ensures responsible use of technology and personal information.
  - Article 43 of the Constitution of Bangladesh – gives citizens the right to privacy.

  It also follows international rules, such as:

  - GDPR (General Data Protection Regulation) – gives users the right to know how their data is used.
  - OECD Privacy Guidelines – global standards for responsible data handling.

- **Ethical AI Compliance**

  VirtualFit uses Artificial Intelligence (AI) to detect skin tones and fit clothes virtually. The project follows the UNESCO AI Ethics Principles, which means the AI is built and used in a fair and honest way. This includes being open about when AI is being used and ensuring the AI doesn't harm or discriminate against anyone. By doing all this, VirtualFit ensures it uses advanced technology responsibly and protects its users.

4. Cultural Impact

- **Fashion Inclusivity**

  VirtualFit is made to support people from different cultures and fashion preferences. The system's catalog can be easily expanded to include not just modern clothes, but also traditional, modest, and indigenous outfits. This helps people express their cultural identity and keeps traditional clothing styles alive in the digital world.

- **Gender Inclusivity**

  VirtualFit is built to be inclusive of all genders and body types. Unlike many fashion platforms that focus only on specific “ideal” body shapes, our catalog features a diverse range of models. The interface is also gender-neutral, meaning it doesn’t assume what kind of clothes someone should wear. This promotes body positivity and helps break down old beauty stereotypes, directly supporting Sustainable Development Goal 5 (Gender Equality).

- **Digital Equity**

  VirtualFit is designed so that anyone with a smartphone or a standard computer can use it. There’s no need for expensive devices or fancy software. This makes it easy for people in rural or underdeveloped areas to use the technology, just like those in big cities. In this way, VirtualFit helps close the digital gap and gives more people access to the benefits of advanced technology.

5.2 Impact of this project on environment and sustainability

1. Environmental Impact

- **Reduction in Reverse Logistics Emissions**

  When customers return clothing purchased online, it creates a significant environmental burden. Each return requires extra packaging, transportation, and reprocessing, all of which contribute to CO_2 emissions. VirtualFIT helps users make more informed purchasing decisions by providing a realistic preview of the garment's fit and style. This can significantly reduce product return rates, which in turn decreases the fuel, packaging, and energy consumed by reverse logistics, making online shopping a more eco-friendly process.

- **Sustainable Digital Infrastructure**

  VirtualFIT is designed to run on modern cloud services like Google Colab, AWS, or GCP. These platforms are engineered for energy efficiency and are increasingly powered by renewable energy sources; for example, Google Cloud operates on 100% renewable energy. By leveraging this elastic cloud infrastructure, our system scales its energy consumption based on demand, avoiding the waste associated with maintaining idle, on-premise servers and promoting a more sustainable digital footprint.

- **Paperless Retail Process**

  By its nature, VirtualFIT supports a fully digital shopping journey. This helps reduce the fashion industry's reliance on physical materials that contribute to waste, such as:

  - Printed product catalogues
  - Paper receipts and return slips
  - Physical flyers and advertisements This shift toward a paperless model helps conserve resources and reduces the environmental impact of printing and physical distribution.

2. Sustainability Goals Alignment (SDG Mapping)

The United Nations Sustainable Development Goals (SDGs) are 17 global goals for a more sustainable future. Our project, VirtualFIT, directly supports several of these goals by using technology to create a more efficient, inclusive, and responsible fashion ecosystem.

- **SDG 3 – Good Health and Well-being**

  VirtualFIT promotes public health by providing a safe, remote alternative to trying on clothes in crowded physical stores. It also supports mental well-being by fostering body positivity and confidence, allowing users to see how clothes look on a diverse range of realistic body types.

- **SDG 5 – Gender Equality**

  Our platform is designed to be inclusive of all users. The interface is gender-neutral, and the model catalog can feature people of all genders and body types, challenging traditional beauty standards and promoting equality in fashion representation.

- **SDG 9 – Industry, Innovation and Infrastructure**

  VirtualFIT is an innovative digital solution for the fashion retail industry. It leverages modern AI and cloud infrastructure to solve a key business challenge, creating new opportunities in fashion technology and contributing to a more resilient and advanced digital economy.

- **SDG 10 – Reduced Inequalities**

  Accessible on any standard smartphone or computer, VirtualFIT helps bridge the digital divide. It gives people in rural or lower-income areas access to the same high-quality styling and shopping tools as those in major urban centers, promoting greater digital equity.

- **SDG 12 – Responsible Consumption and Production**

  The most direct impact of our project is on this goal. By helping users make more accurate purchasing decisions, VirtualFIT helps reduce the number of unnecessary clothing returns, which cuts down on waste, shipping emissions, and excess packaging, fostering more responsible consumption patterns.

- **SDG 13 – Climate Action**

  By reducing the volume of returns in the e-commerce supply chain, VirtualFIT helps to lower the overall carbon footprint of the fashion industry, directly supporting climate action.

- **SDG 17 – Partnerships for the Goals**

  VirtualFIT is a project that brings together multiple fields: academic research in AI, cloud infrastructure from tech companies, the business needs of fashion brands, and the end-user. This kind of collaboration supports SDG 17 by demonstrating how partnerships can create positive global change.

Chapter 6 - Project Planning and Budget

6.1 Project Planning

The planning for the VirtualFIT project was organized into five distinct phases, moving from initial research and strategy to final system integration and documentation. The project's timeline was dynamic, with several phases overlapping to ensure efficient progress. Key strategic pivots, such as moving from unstable diffusion models to the more robust VITON-HD framework, were critical to the project's success. The detailed timeline of all tasks is visualized in the Gantt chart below.

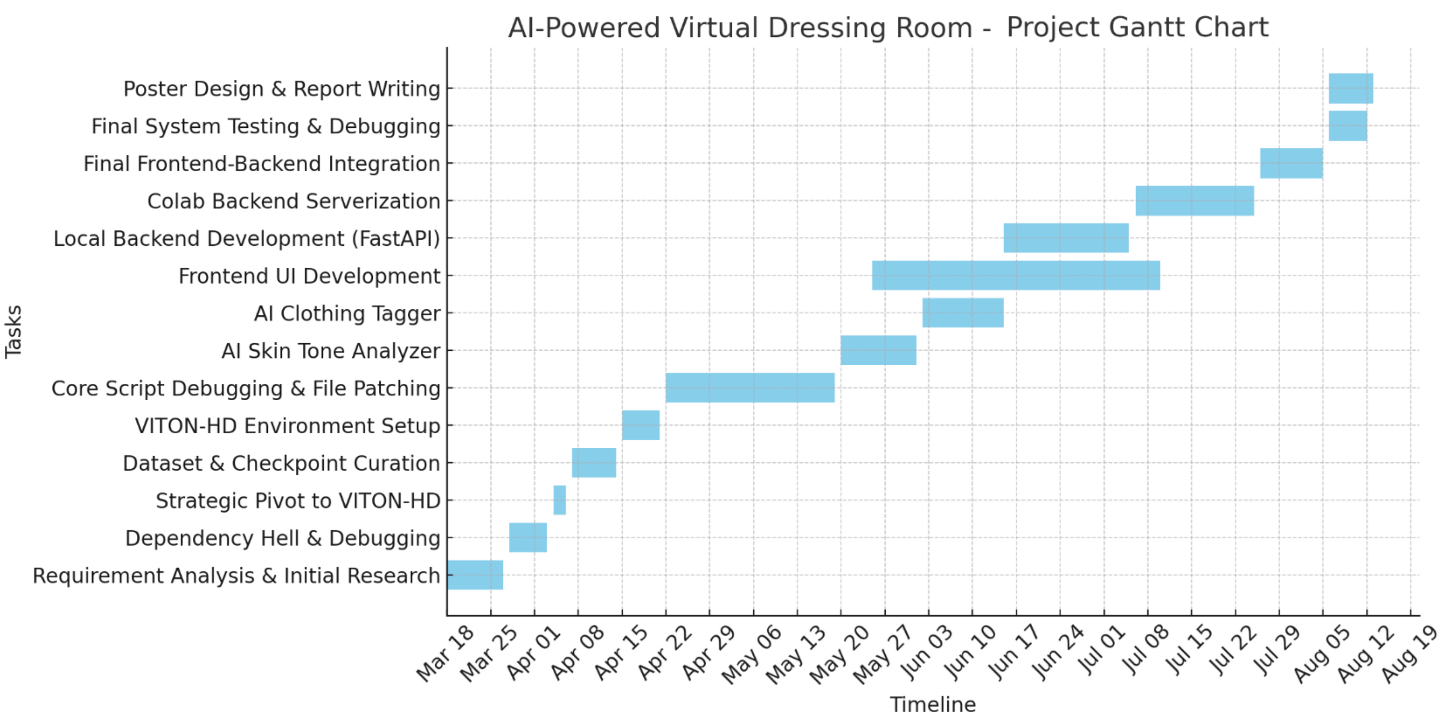

Figure 6-1: A Gantt chart visualizing the project timeline with overlapping tasks

6.2 Budget

The budget for this project was managed to be cost-effective by leveraging cloud computing resources and open-source software. The primary costs were associated with GPU access for model training and inference, data storage, and essential hardware for development.

Item	**Description**	**Approximate Cost (BDT)**
Cloud Computing	Google Colab Pro subscription for high-performance GPU access.	2,000
Cloud Storage	6-month Google One subscription for storing large datasets and model checkpoints.	3,500
Hardware	A high-quality webcam for testing the "Match My Skin Tone" feature.	5,000
Networking	A USB WiFi adapter to ensure a stable connection for development.	1,000
Miscellaneous	Contingency for additional software tools or unforeseen expenses.	2,000
Total		**13,500**

Figure 6.2: The approximate budget for the VirtualFIT project.

Budget Justification

The budget for the VirtualFIT project was carefully planned to maximize the use of cost-effective cloud resources and essential hardware. The total cost reflects the necessary investments to develop a high-performance AI pipeline and a fully functional web application.

1. Google Colab Pro

- **Purpose:** To access the high-performance GPUs required for running the VITON-HD deep learning model.
- **Benefit:** Using a professional cloud GPU service like Colab Pro was essential for the project. It provided the necessary computational power to run the complex generative models, significantly reducing the time required for inference and testing.
- **Cost:**
  - Subscription: ৳2,000

2. Data Storage (Google One)

- **Purpose:** To securely store and access the large, multi-gigabyte VITON-HD dataset, model checkpoints, and generated results.
- **Benefit:** A Google One subscription provided a reliable and fast cloud storage solution that integrates seamlessly with Google Colab, which was crucial for our data-centric workflow.
- **Cost:**
  - 6-Month Subscription: ৳3,500

3. Webcam

- **Purpose:** To develop and test the "Match My Skin Tone" feature, which requires capturing a live image of the user.
- **Benefit:** A high-quality webcam was a necessary investment to ensure that the skin tone detection algorithm received a clear, well-lit image, leading to more accurate and reliable results for the user.
- **Cost:**
  - Webcam Purchase: ৳5,000

4. WiFi Adapter

- **Purpose:** To ensure a stable, high-speed internet connection during the development process.
- **Benefit:** A reliable connection was critical for managing large file transfers between our local machines and Google Drive, as well as for uninterrupted access to the Colab environment, which prevented delays and data corruption.
- **Cost:**
  - WiFi Adapter Purchase: ৳1,000

5. Miscellaneous

- **Purpose:** To provide a contingency fund for any unforeseen expenses.
- **Benefit:** This buffer was allocated for potential costs such as small software subscriptions, necessary plugins, or any unexpected technical requirements that could have arisen during the development and deployment phases.
- **Cost:**
  - Contingency: ৳2,000

Summary of Total Costs

1. Google Colab Pro: ৳2,000
2. Data Storage: ৳3,500
3. Webcam: ৳5,000
4. WiFi Adapter: ৳1,000
5. Miscellaneous: ৳2,000

Total: ৳13,500

In conclusion, the total cost for the project is approximately **৳13,500**. This budget covers all the essential resources, including cloud GPU access for the AI model, secure data storage for our gbhlarge dataset, and the necessary hardware for developing and testing our innovative user-facing features, ensuring the successful completion of the VirtualFIT application.

Chapter 7 - Complex Engineering Problems and Activities

7.1 Complex Engineering Problems (CEP)

The development of VirtualFIT required solving a series of complex engineering problems. The challenges went far beyond simple software implementation, demanding in-depth analysis, strategic decision-making, and the integration of multiple, often conflicting, technical systems. The table below maps the project's challenges to the defined Complex Engineering Problem attributes.

Attributes		Addressing the complex engineering problems (P) in the project
P1	Depth of knowledge required (K3-K8)	The project required deep knowledge of Computer Vision (K3), Generative AI models like GANs (K4), Software Engineering for system design (K5), and modern web development tools (K6). A significant portion of the work involved researching and understanding multiple academic papers and GitHub repositories (K8) to debug and implement the VITON-HD model.
P2	Range of conflicting requirements	A central conflict was Stability vs. Quality. Newer diffusion models promised higher quality but were unstable in our cloud environment. The older VITON-HD model was stable but required significant engineering to achieve a good result. We also had to balance User Experience (fast, seamless) with Technical Feasibility (pre-generating results vs. slow real-time inference).
P3	Depth of analysis required	There was no single, obvious solution. We had to analyze multiple AI models and diagnose complex "dependency hell" issues between libraries like PyTorch, JAX, and OpenCV. This required abstract thinking to design a new, functional pipeline by patching and combining different software components that were not originally designed to work together.
P7	Interdependence	The system is highly interdependent. The final image quality is directly dependent on the AI generator, which is dependent on the cloth warping module, which in turn depends on a perfect set of seven pre-processed input files (pose, parse map, agnostic image, etc.). A failure or low-quality output in any one of these sub-problems causes a catastrophic failure in the final result.

Table 7.1: A Sample Complex Engineering Problem Attributes

7.2 Complex Engineering Activities (CEA)

The execution of the VirtualFIT project involved a range of complex engineering activities, from managing diverse resources to innovating a practical solution from theoretical research. The table below maps our project activities to the defined attributes.

Attributes		Addressing the complex engineering activities (A) in the project
A1	Range of resources	The project involved diverse resources, including: Human Resources (a team with specialized backend, frontend, and AI roles), Equipment (local PCs for development, high-performance GPUs via Google Colab for training/inference), Information (multiple research papers, GitHub repositories), and Technologies (Python, React, FastAPI, PyTorch, OpenCV, CLIP, ngrok).
A2	Level of interactions	A significant challenge was managing the interactions between different software systems. We had to resolve numerous technical conflicts between the React frontend, the local Python backend (for skin tone analysis), the remote AI backend (running on Colab), and the data storage on Google Drive. This required careful API design and debugging across multiple platforms.
A3	Innovation	The project's innovation was not in creating a new AI model, but in the engineering required to make a complex research model functional and useful. We innovated by creating a stable pipeline for an unstable model and by designing novel, user-facing "Virtual Stylist" features (like the AI Clothing Tagger and Skin Tone Matcher) that make the technology practical.
A4	Consequences to society / Environment	The project has significant positive consequences. Societally, it offers a solution to create a more confident and inclusive online shopping experience. Environmentally, by providing a more accurate preview of an item, this technology can help reduce product returns, which in turn lowers the carbon footprint and waste associated with reverse logistics in the fashion industry.
A5	Familiarity	The project constantly extended beyond our previous experience. We were not just following a tutorial; we were debugging undocumented research code, analyzing cryptic error messages, and designing a multi-part, distributed system. This required applying fundamental engineering principles to novel problems, such as reverse-engineering data formats and patching external libraries to ensure compatibility.

Table 7.2: A Sample Complex Engineering Problem Activities

Chapter 8 - Conclusions

8.1 Summary

This project successfully engineered a stable, end-to-end pipeline for the VITON-HD virtual try-on model, transforming a complex research framework into a functional, user-centric application. The core AI engine is fully implemented, and we have demonstrated the ability to take any clothing item from our catalog and realistically render it onto any of our pre-processed digital models. The most significant achievement of this project was the strategic pivot from a highly ambitious but unreliable automated pre-processing pipeline for raw user images to a more controlled, data-centric model. This decision to prioritize stability by using a high-quality, pre-processed dataset was critical, as it guaranteed a flawless and realistic output for every try-on. This robust technical foundation now serves as the platform for our "VirtualFIT" application and its innovative "Virtual Stylist" features.

8.2 Limitations

While the project achieved its primary goals, it is important to acknowledge its limitations. The most significant of these is the system's reliance on a pre-processed, curated dataset of models. Our experiments confirmed that the VITON-HD model, like many GANs, is highly sensitive to the quality and pixel-perfect consistency of its seven required input files. Our attempts to build a fully automated pipeline for raw, "in-the-wild" user photos produced distorted results due to minor but compounding inconsistencies between the outputs of the different AI models used for parsing and pose estimation. This restricts the current application to our high-quality catalog and prevents a fully personalized try-on experience for any user.

Furthermore, the dataset itself presents a limitation. The zalando-hd-resized dataset, while high-quality, consists predominantly of female models wearing Western-style tops. This narrows the scope of the try-on experience and does not yet represent the full diversity of body types, genders, and cultural attire that a global fashion platform would require.

8.3 Future Improvement

The successful implementation of the core AI pipeline opens up several exciting and commercially viable avenues for future work.

The most significant future improvement would be to solve the challenge of processing raw user photos. With further research and funding, we would aim to integrate a state-of-the-art, multi-class human parsing model, such as Graphonomy or SCHP. A more advanced parsing engine could potentially generate a "blueprint" of a user's body with the precision and consistency required by the VITON-HD generator. Achieving this would be a major breakthrough, transforming the application from a B2B catalog tool into a fully personalized B2C experience and eliminating the need for professional model photoshoots for our clients.

A second area for improvement would be to expand the system's capabilities beyond tops. We would work on integrating models that can handle different garment types, such as trousers, skirts, and dresses. This would involve curating new datasets and potentially fine-tuning the GMM and ALIAS Generator to better understand the unique warping and draping physics of different clothing items, leading to a more comprehensive and versatile virtual fitting room.

References

1. Attallah, K., Zaky, G., Abdelrhim, N., Botros, K., Dife, A., & Negied, N. (2024). A cost-efficient approach for creating virtual fitting room using generative adversarial net orks (GANs). International Journal of Advanced Computer Science and Applications (IJACSA), 15(1). https://doi.org/10.48550/arXiv.2402.00994

2. Gill, S., & Brownbridge, K. (n.d.). The truth of standard sizing. Retrieved from

https://e-space.mmu.ac.uk/611547/2/The%20truth%20of%20standard%20sizing.pdf

3. Su, J., Wang, C., Ma, X., Zeng, W., & Wang, Y. (2022). VirtualPose: Learning generalizable 3D human pose models from virtual data. In Computer Vision – ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part VI (pp. 55-71). https://doi.org/10.1007/978-3-031-20068-7_4

4. Idrees, S., Gill, S., & Vignali, G. (2023). Mobile 3D body scanning applications: A revie of contact-free AI body measuring solutions for apparel. International Journal of Fashion Design, Technology and Education, 1161-1172. https://doi.org/10.1080/00405000.2023.2216099

5. Cimolin, V., Paraskevopoulos, I. T., Sala, M., Tarabini, M., & Galli, M. (2022). The smart body concept as a demonstration of the overarching utility and benefits of 3D avatars in retail, health, and wellbeing: An accuracy study of body measures from 3D reconstruction. Multimedia Tools and Applications, 82(7), 11079–11098. https://doi.org/10.1007/s11042-022-13661-x

6. Han, X., Wu, Z., Wu, Z., Yu, R., & Davis, L. S. (2019). "VITON: An image-based virtual try-on network." IEEE Conference on Computer Vision and Pattern Recognition, 7543-7552.

7. Barroso-Laguna, A., Munukutla, S., Prisacariu, V. A., & Brachmann, E. (2024). Matching 2D images in 3D: Metric relative pose from metric correspondences. arXiv:2404.06337v1 [cs.CV]. https://arxiv.org/html/2404.06337v1

8. Liu, Y., Liu, Y., Xu, S., Cheng, K., Masuko, S., & Tanaka, J. (2020). Comparing VR- and AR-based try-on systems using personalized avatars. Electronics, 9(11), 1814. https://doi.org/10.3390/electronics9111814

9. Lee, H., Xu, Y., & Porterfield, A. (2021). Antecedents and moderators of consumer adoption toward AR-enhanced virtual try-on technology: A stimulus-organism-response approach. International Journal of Consumer Studies. https://doi.org/10.1111/ijcs.12760

10. Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., & Sheikh, Y. (2018). OpenPose: Realtime multi-person 2D pose estimation using part affinity fields. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. XXX, no. XXX, 1. https://arxiv.org/pdf/1812.08008

11. Kim, J.-W., Choi, J.-Y., Ha, E.-J., & Choi, J.-H. (2023). Human pose estimation using MediaPipe Pose and optimization method based on a humanoid model. Applied Sciences, 13(4), 2700. https://doi.org/10.3390/app13042700

12. J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks," arXiv preprint arXiv:1703.10593, 2017, revised Aug. 2020. [Online]. https://arxiv.org/pdf/1703.10593

13. H. Wang, Z. Zhang, D. Di, S. Zhang, and W. Zuo, "MV-VTON: Multi-View Virtual Try-On with Diffusion Models," arXiv preprint arXiv:2404.17364, 2024. [Online]. Available: https://arxiv.org/abs/2404.17364

14. K. Sun, J. Cao, Q. Wang, L. Tian, X. Zhang, L. Zhuo, B. Zhang, L. Bo, W. Zhou, W. Zhang, and D. Gao, "OutfitAnyone: Ultra-high Quality Virtual Try-On for Any Clothing and Any Person," arXiv preprint arXiv:2407.16224, 2024. [Online]. Available: https://arxiv.org/abs/2407.16224

15. S. Ning, D. Wang, Y. Qin, Z. Jin, B. Wang, and X. Han, "PICTURE: PhotorealistIC Virtual Try-on from UnconstRained dEsigns," arXiv preprint arXiv:2312.04534, 2023. [Online]. Available: https://arxiv.org/abs/2312.04534

16. K. Li, J. Zhang, S.-Y. Chang, and D. Forsyth, "Wearing the Same Outfit in Different Ways -- A Controllable Virtual Try-on Method," arXiv preprint arXiv:2211.16989, 2022. [Online]. Available: https://arxiv.org/abs/2211.16989

17. S. Choi, S. Park, M. Lee, and J. Choo, "VITON-HD: High-Resolution Virtual Try-On via Misalignment-Aware Normalization," arXiv preprint arXiv:2103.16874, 2021. [Online]. Available: https://arxiv.org/pdf/2103.16874

18. X. Dong, F. Zhao, Z. Xie, X. Zhang, D. K. Du, M. Zheng, X. Long, X. Liang, and J. Yang, "Dressing in the Wild by Watching Dance Videos," arXiv preprint arXiv:2203.15320, 2022. [Online]. Available: https://arxiv.org/abs/2203.15320

19. C.-Y. Chen, L. Lo, P.-J. Huang, H.-H. Shuai, and W.-H. Cheng, "FashionMirror:

Co-attention Feature-remapping Virtual Try-on with Sequential Template Poses," National Chiao Tung University, Hsinchu, Taiwan, 2021. [Online]. Available: https://openaccess.thecvf.com/content/ICCV2021/papers/Chen_FashionMirror_Co-Attention_Feature-Remapping_Virtual_Try-On_With_Sequential_Template_Poses_ICCV_2021_paper.pdf

20. K. Ayush, S. Jandial, A. Chopra, and B. Krishnamurthy, "Powering Virtual Try-On via Auxiliary Human Segmentation Learning," in Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), 2019. [Online]. Available: https://openaccess.thecvf.com/content_ICCVW_2019/papers/CVFAD/Ayush_Powering_Virtual_Try-On_via_Auxiliary_Human_Segmentation_Learning_ICCVW_2019_paper.pdf

21. X. Han, X. Hu, W. Huang, and M. R. Scott, "ClothFlow: A Flow-Based Model for Clothed Person Generation," Malong Technologies, Shenzhen, China, 2020. [Online]. Available: https://openaccess.thecvf.com/content_ICCV_2019/papers/Han_ClothFlow_A_Flow-Based_Model_for_Clothed_Person_Generation_ICCV_2019_paper.pdf

22. D. Roy, S. Santra, and B. Chanda, "LGVTON: A Landmark Guided Approach to Virtual Try-On," arXiv preprint arXiv:2004.00562, 2020. [Online]. Available: https://arxiv.org/abs/2004.00562v1