VIRTUAL TRY-ON SYSTEM

Experience e prova like never before, transforming online shopping into a personalized adventure.

SUPERVISUR : AHMED HESHAM

EPROVA

جامعة حلوان
Helwan University
1975

Team members :

- Marwan shamel 20210895
- Marwan abdelrady 20210896
- Marwan magdy 20210898
- Mariam samy 20210913
- Mariam fathy 20210921
- Mariam elsayed 20210908

agenda

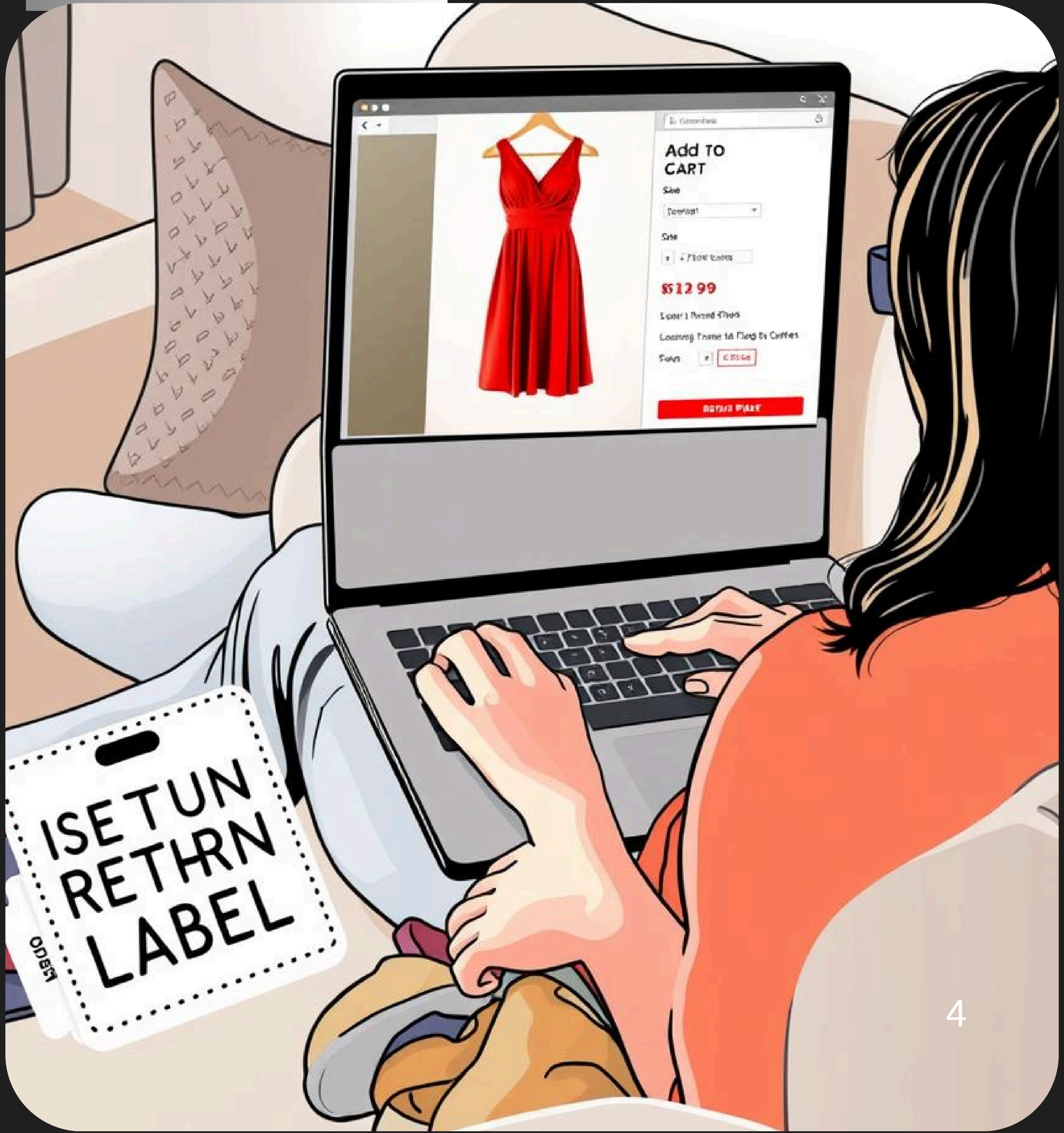
1. Problem Statement
2. Solution
3. Related Work
4. system work flow
5. Challenges
6. Experiments
7. implemtation
8. Evalution
9. future work
10. Conclusion

EPROVA

the problem

Online clothes shopping offers convenience and variety, but a key issue remains: users can't try on clothes before buying. This causes uncertainty, poor fit, and high return rates. Many customers hesitate to purchase due to doubts about size and style.

Most platforms lack interactive tools to visualize clothes on real users, relying instead on static images and models. This makes the experience harder for first-time shoppers or those with limited fashion knowledge—especially in areas without access to physical stores.



solution

E-Prova aims to solve these problems by introducing a smart virtual try-on feature that allows users to see how clothes would look on their own photos in a realistic and interactive way. The system also provides all core e-commerce functionalities and aims to make online fashion shopping more personalized, confident, and satisfying for every user.

Related Work

1-Image-Based Virtual Try-On using GANs:

- Deform (warp) the garment to fit the person's body shape.
- Use GANs to generate the final image.
- Limitation: Poor generalization to complex backgrounds and diverse poses.

2-Diffusion-Based Virtual Try-On:

- Apply diffusion models like TryOnDiffusion with dual UNet architecture.
- Some methods treat the task as image inpainting using reference images.
- Limitation: Difficulty in preserving fine garment details.

Related Work (cont.)

3-Conditional Control in Diffusion Models:

- Enhance generation by adding controls like pose, edges, depth.
- Models like ControlNet and IP-Adapter allow more accurate generation using both text and image prompts.

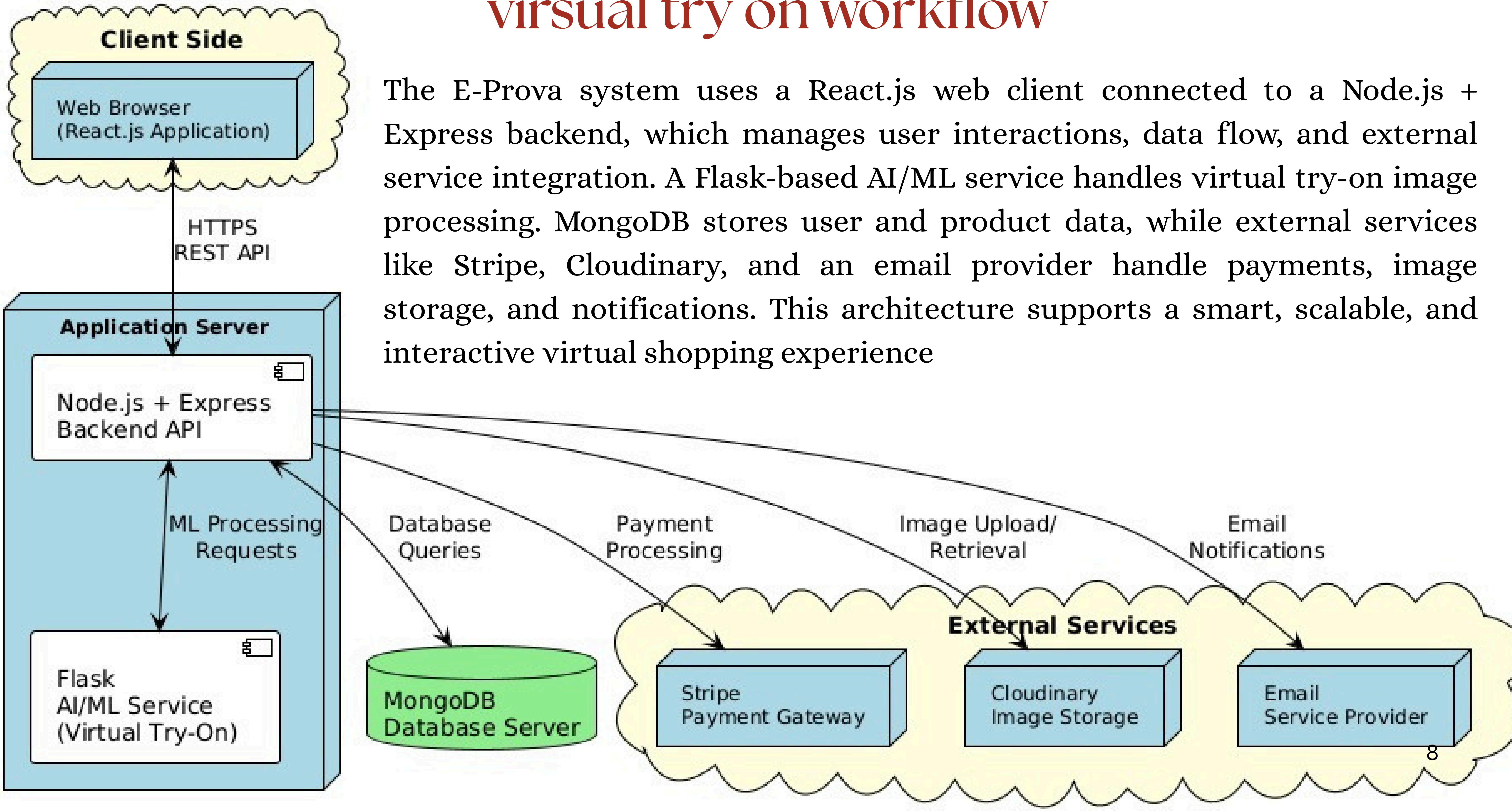
4-Customization of Diffusion Models:

- Fine-tune models using a few personal images (few-shot learning).
- Achieve better adaptation to real-world data without catastrophic forgetting.
- Also useful for tasks like image completion or restoration.

EPROVA

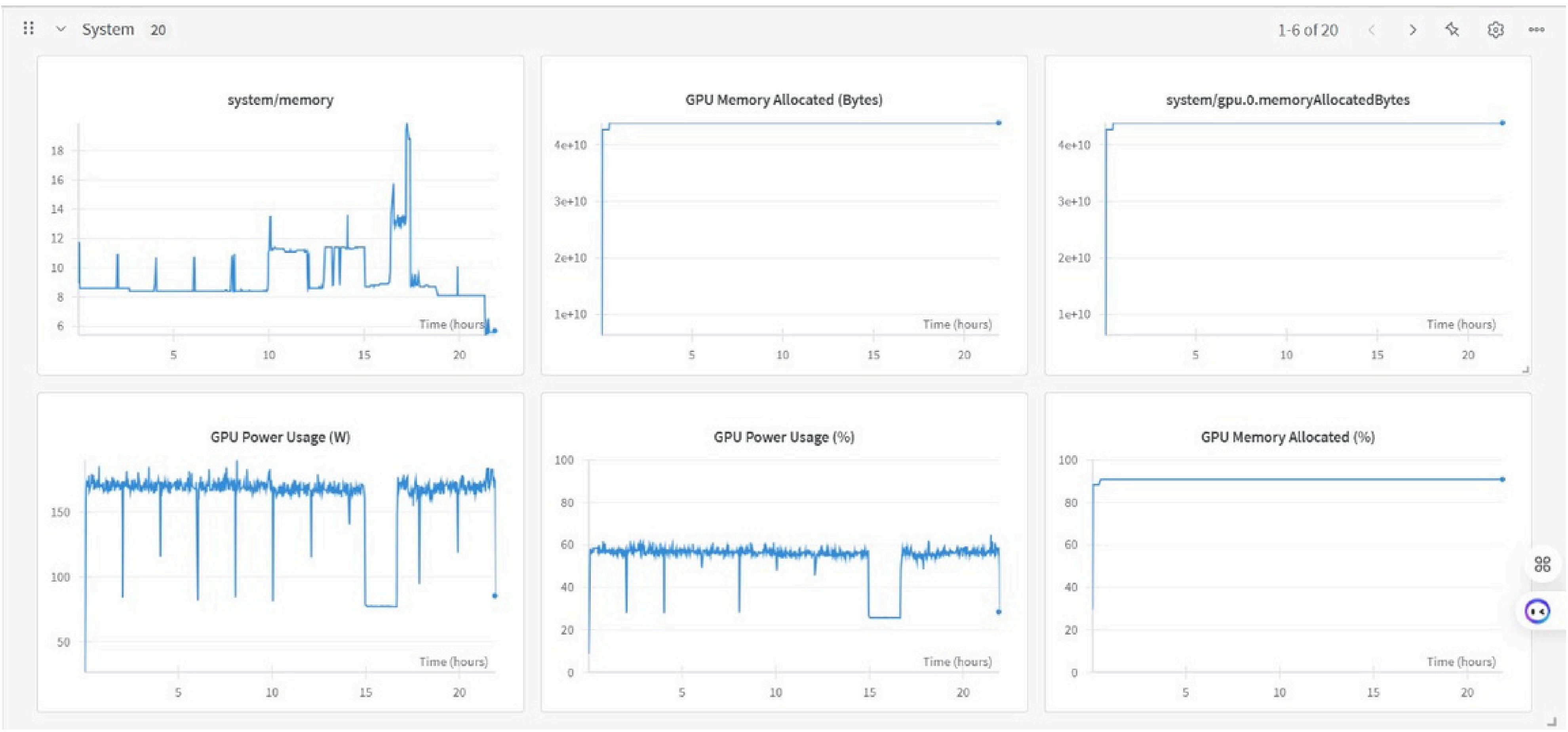
virsual try on workflow

The E-Prova system uses a React.js web client connected to a Node.js + Express backend, which manages user interactions, data flow, and external service integration. A Flask-based AI/ML service handles virtual try-on image processing. MongoDB stores user and product data, while external services like Stripe, Cloudinary, and an email provider handle payments, image storage, and notifications. This architecture supports a smart, scalable, and interactive virtual shopping experience



Challenges

1. Limited Computational Resources
    - Problem:
        - Only 4GB GPU memory available (Colab/Kaggle free tier)
        - Required 12GB+ for full IDM-VTON training
    - Impact:
        - 83% of experimental runs crashed during diffusion refinement
        - 3-4 hour processing times per image on CPU fallback



2. Prohibitive Cloud Costs
    - Problem:
        - Cloud GPUs (A100/V100) cost $2.1-$4.8/hour
        - Full training required 200+ GPU hours
    - impact :
        - Budget constraints forced 75% reduction in training iterations
        - Reliance on pre-trained checkpoints with limited customization

Challenges

3. Session Instability

- Problem:
  - Runtime disconnections after 12h (Colab limits)
  - Memory leaks during TPS warping
- Impact:
  - Lost 17 completed experiments due to session crashes
  - Required manual checkpointing every 30 minutes

Mitigation Strategies

- Adopted mixed-precision (FP16) training → 40% memory reduction
- Implemented gradient checkpointing → Enabled larger batch sizes
- Used Hugging Face cached models → Avoided redundant downloads

Experiment 1: Lady VITON

Goal : Test a baseline virtual try-on system

Setup: Used Lady VITON repo with VITON-HD dataset

Preprocessing: No automatic preprocessing – required manual alignment

Results: Output lacked realism and consistencyand and Clothes were poorly aligned with user body.

Experiment 2: IDM-VTON

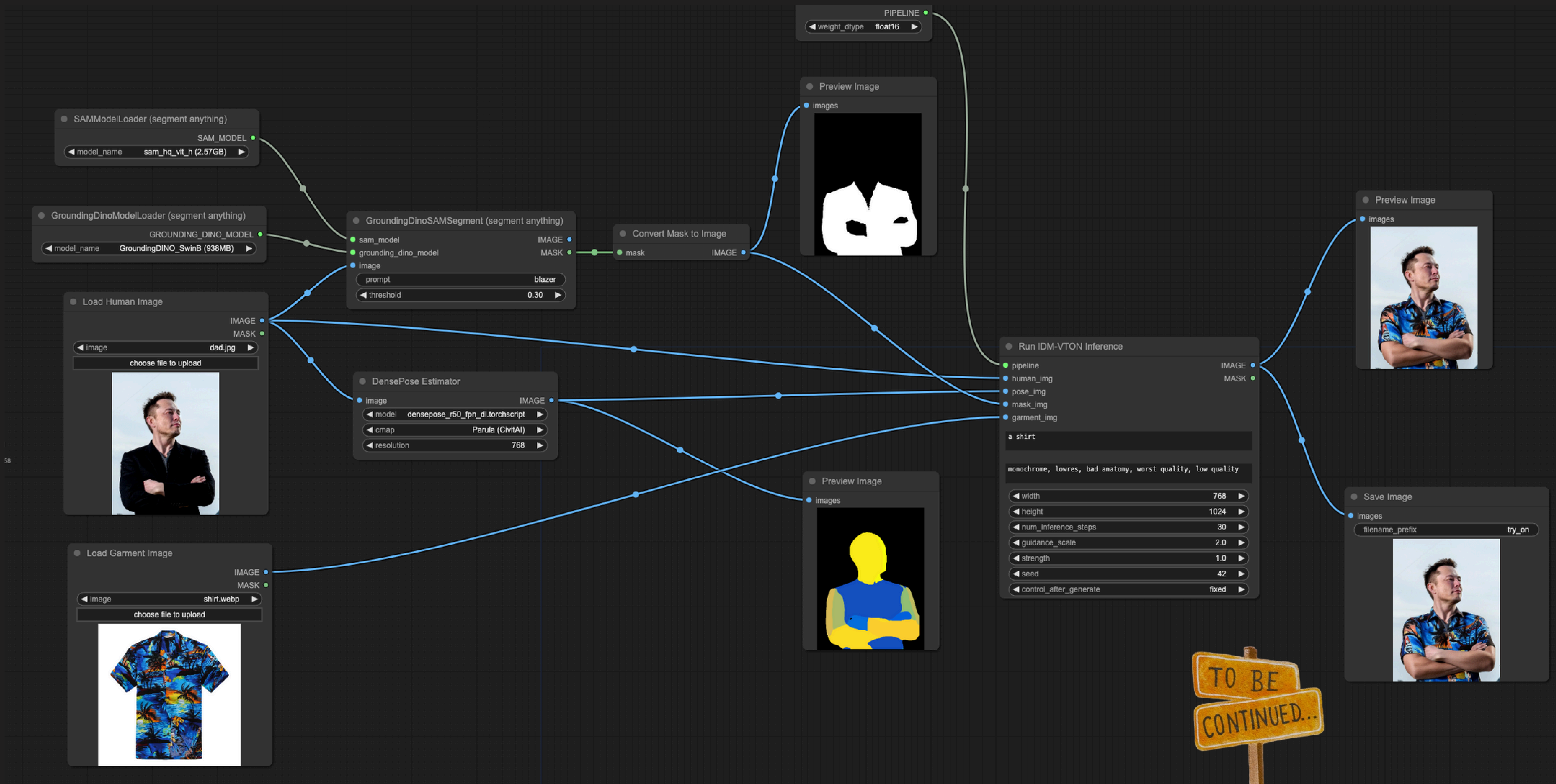
Goal : Improve realism and automate the pipeline

Setup: Used IDM-VTON with the same dataset

Preprocessing: Better fit and more realistic overlay,Handled occlusions and body shapes smoothly

Results: Better fit and more realistic overlay and Handled occlusions and body shapes smoothly.

implementation flow



implementation flow (cont)

1. Load Human Image

The first thing you do is upload an image of the person (the person you will try the clothing on)

2. Load Garment Image

Then you upload an image of the clothing you want to try on (the shirt or blouse, for example)

3. Segment the Garment (Cloth Mask Generation)

- Using:GroundingDINO + SAM (Segment Anything)
- The clothing in the image is automatically identified based on a prompt word like "blazer."
- SAM creates a mask to define the boundaries of the clothing.
- This mask is converted to a black and white image (binary mask).

implementation flow (cont)

4. Get Human PoseUsing:

- DensePose Estimator
- This extracts the body shape and distribution of parts (head, arms, body) as a pose
- image.This image helps in adjusting the clothing to fit the person's body

5. Run IDM-VTON InferenceIt

takes as inputs :human_img → the person's image
garment_img → the clothing image
mask_img → the mask that defines the clothing
pose_img → the image showing body distribution
garment_des → a simple description of the clothing (like "a shirt")
the output:A new image of the person as if they are wearing the clothing you selected.

6. Output + Save ImageIt displays the output as a new image that you can save.

Evaluation

```
100% 64/64 [00:08<00:00,  7.43it/s]
{'ssim_score': 0.7334049344062805, 'lpips_score': 0.25036758184432983, 'fid_score': 35.53354407692765
Steps: 100% 250/250 [51:45<00:00, 12.42s/it, lr=5e-6, step_loss=0.00806]
```

this is like a comparison between ladi and IDM in :

SSIM (Structural Similarity Index): Measures image similarity

LPIPS (Learned Perceptual Image Patch Similarity): A perceptual metric where lower is better

Metric	LaDI-VTON	IDM-VTON	Best
SSIM ↑	0.872	0.870	LaDI-VTON
LPIPS ↓	0.156	0.102	IDM-VTON
FID ↓	8.85	6.29	IDM-VTON
CLIP-I ↑	0.834	0.883	IDM-VTON

FID (Fréchet Inception Distance): Lower FID indicates better alignment with real distribution

CLIP-I (CLIP Similarity Score): Evaluates how well the generated image matches the textual description (or reference image) using CLIP embeddings

Why Better?

- **Automatic Preprocessing:**

handles preprocessing steps (like pose estimation and parsing) automatically

- **Better Garment Fit:**

aligns clothes more accurately on the body, especially with complex poses.

- **More Realistic Results:**

The output images are more natural and preserve garment details better.

- **Stronger Performance:**

works better with complex backgrounds and diverse user poses.

- **Easier Integration:**

IDM-VTON is modular and easier to use in a complete system.

Future work :

- Add 3D body modeling for more accurate and multi-angle try-on.
- Enable real-time try-on using webcam and AR.
- Implement AI-based size recommendations.
- Use user feedback to refine model accuracy.
- Develop a mobile app for better accessibility.

EPROVA

Conclusion

E-Prova combines AI and e-commerce to enhance online fashion shopping by letting users try on clothes virtually using their own photos. Using the IDM-VTON model, it delivers realistic and well-aligned try-on results. The platform also includes essential shopping features, offering a personalized and user-friendly experience.

Frontend repo

E-PROVA

Notebook IDM

Backend repo

Notebook LADI

Thank You ! ☺