

# I Like Your Style

I Scanned, Magic Happened

December 8th 2023



# Team Introduction

**Victor Tennenoni:** VP at Eurazeo (\$34b AUM VC Fund), Bachelor in CS from 42 NUS

**Mashaal Musleh:** 5th year Ph.D candidate at University of Minnesota

**Grace Fan:** 4th year PhD candidate at Northeastern University, studying data management.

**Brian Formento:** 4th year Phd Candidate at NUS, studying adversarial robustness in NLP.

**Hantian Zhang:** 5th year PhD candidate at Georgia Tech, working on responsible data management



# I Like Your Style



Scan QR

I Scanned, Magic Happened

I See It, I Want It

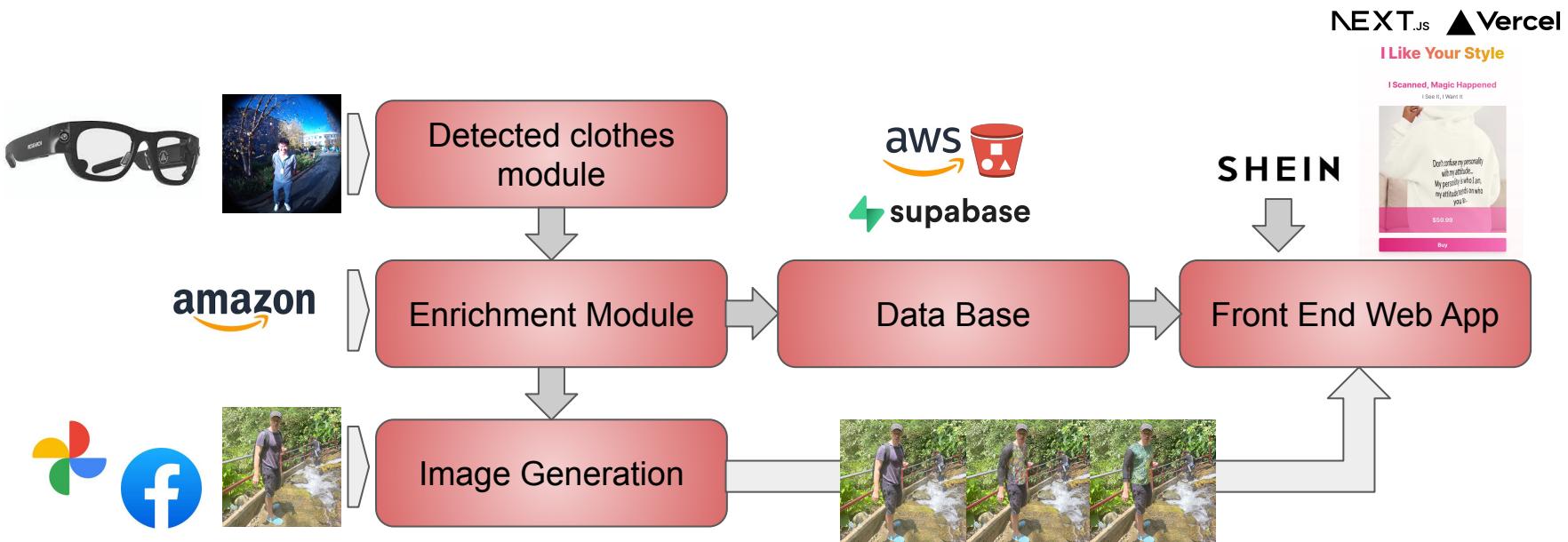
A product advertisement featuring a person wearing a white turtleneck sweater. The sweater has a quote printed on it: "Don't confuse my personality with my attitude... My personality is who I am, my attitude depends on who you are". Below the sweater is a pink price tag with the price \$59.99. At the bottom is a pink button with the word "Buy".

# Revolutionizing Interaction: The Aria Glasses Experience

- 1. Integrate AR and AI:** Combining Augmented Reality with Artificial Intelligence, including object detection, narration models, and control net for advanced, real-time interactions.
- 2. Maximize Aria Glasses Tech:** Utilizing Aria Glasses' high-resolution cameras, precise eye detection, portability, and support for rich video formats to deliver an unparalleled technical experience.
- 3. Developed Full End-to-End User Experience:** Crafted a complete, seamless user journey, enhancing interaction and engagement, designed for a young and diverse audience.
- 4. Ethical Model Usage:** Conscientiously address privacy and bias risks in model development and application.



# Architecture



# Integrate AR and AI: Base Case

Directly walk up to someone and say:

“I like your style”

**Extracted clothing description:**

Color: Grey

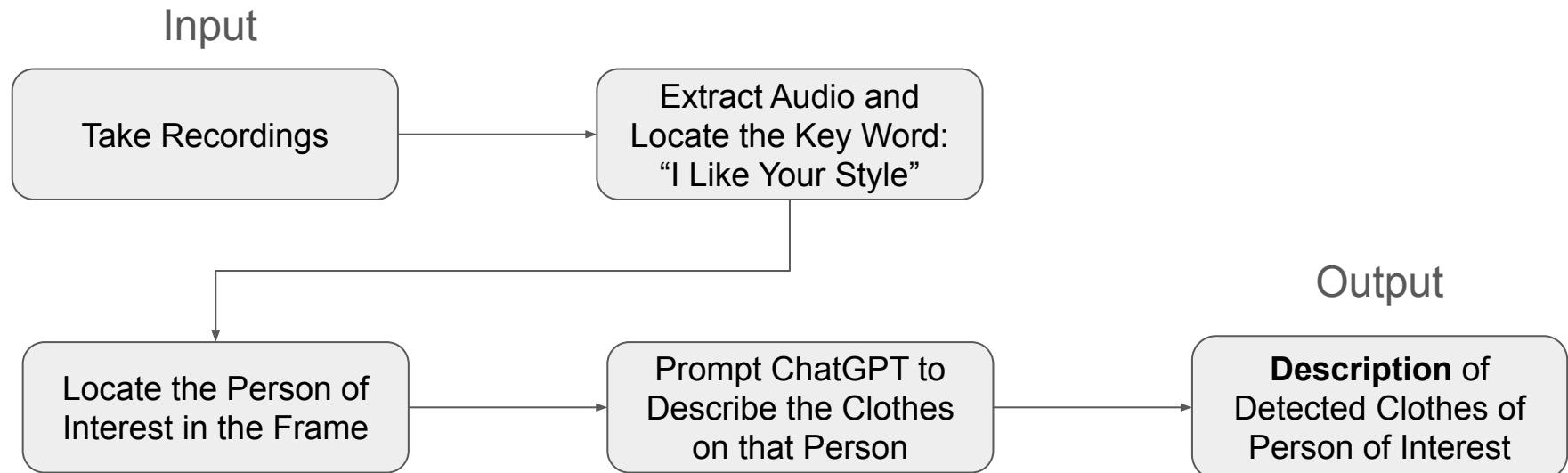
Fabric: Appears to be a synthetic blend

Length: Long-sleeve

Pattern: Solid with some color blocking (green and light blue details)



# Integrate AR and AI: Detect & Describe



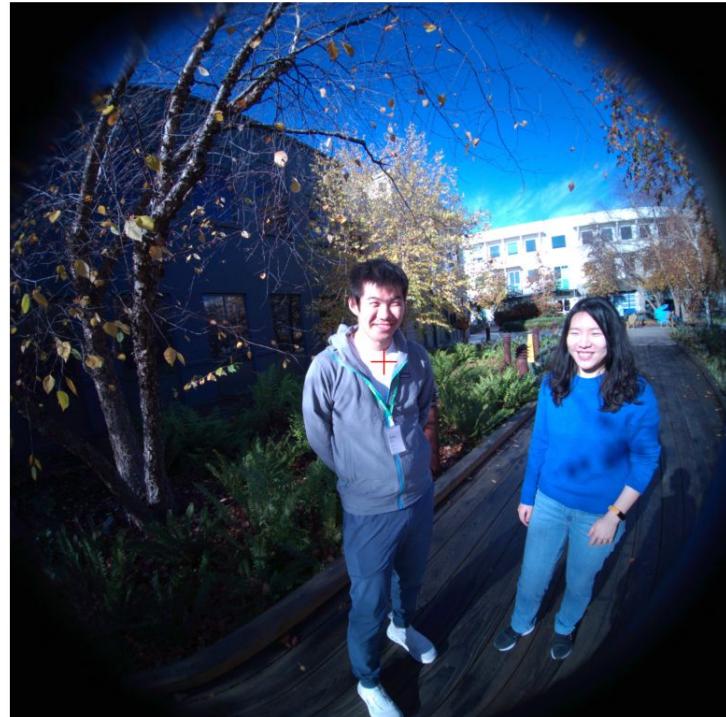
# Integrate AR and AI Upgrade #1: Eye Tracking

Walk up to 2 people and say to one person:

“I like your style”

**Extracted clothing description:**

Color: Blue/grey, Fabric: Synthetic blend, Length: Long-sleeve, Pattern: Solid



# Integrate AR and AI Upgrade #2: Crowds

Walk up to a crowd and say, to one person,  
“I like your style”

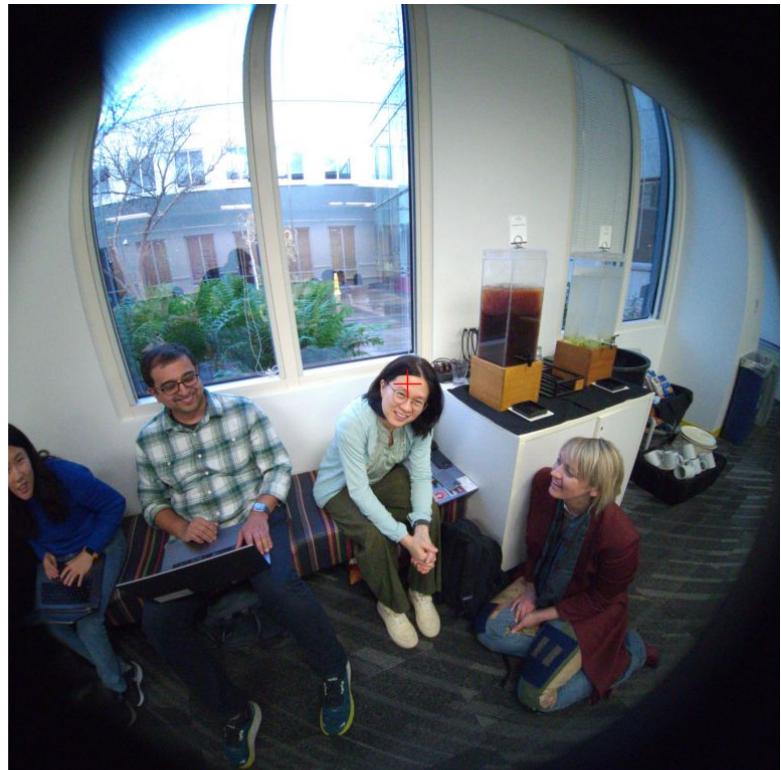
**Extracted clothing description:**

Color: Light Blue

Fabric: Knit (appears to be)

Length: Long sleeves

Pattern: Solid



# Integrate AR and AI Upgrade #3: Peripheral View

Walk up to a crowd and say, to one person,  
“I like your style”

**Extracted clothing description:**

Color: Green

Fabric: It appears to be a sweater fabric, likely a knit but the exact fabric content cannot be determined from the image

Length: Long-sleeved

Pattern: Graphic pattern (with what seems to be a depiction of a character or similar graphic on the front).



# Findings

- Combining speech detection and eye tracking leads to high accuracy when locating person of interest
- Eye tracking helps GPT focus on the person of interest and only generate their clothing description

# Enrichment Module

Integrate AR and AI & Full End-to-End User Experience

- Goal: Enrich (customize) the detected cloths using personalized timeline.
- Examples:
  - “white T-shirt with graphic” ->
    - “Calvin Klein white T-shirt with graphic” or
    - “T-shirt with california style graphic”
  - “blue skirt” ->
    - “knee-length blue skirt”
- How:
  - Input: Purchase History
  - Generate **summary of style/preferences** [ChatGPT]
  - Personalized search term based on **detected cloth & style summary** [Chat GPT]
- Challenges:
  - Vague abstract terms
  - Quality of “SheIn” search APIs
  - Multimodal AI? Items in embedding space?
  - Can preference make different style?



# Integrate AR and AI: ControlNet

What we tried

- Image generation to visualize timeline

What we did

- Implemented:
  - Inpainter + Refiner
  - ControlNet

What we learned

- How to integrate an entire pipeline (from Aria & Timeline to front end)



## Your personal timeline

# Maximise Aria and Timeline Capabilities

### Video Interaction with friends

Sun June 6 2023 12:57:14 GMT+0800 (Singapore Standard Time)



[More details](#)

### Purchases you made



Black top

White Shoes

Red Mask

Date Range

05/14/2023 - 12/06/2023

All Events

### Hanging out in a waterfall

Sun May 14 2023 12:57:14 GMT+0800 (Singapore Standard Time)



[More details](#)

### Welcome to TL-QA

TL-QA \$ What would I look like with my friend's clothes? I interacted with him on the 6th of June. You can take my personal Amazon purchases to give clothes recommendations with Shein. Use my Picture from East Java to showcase what I would look like

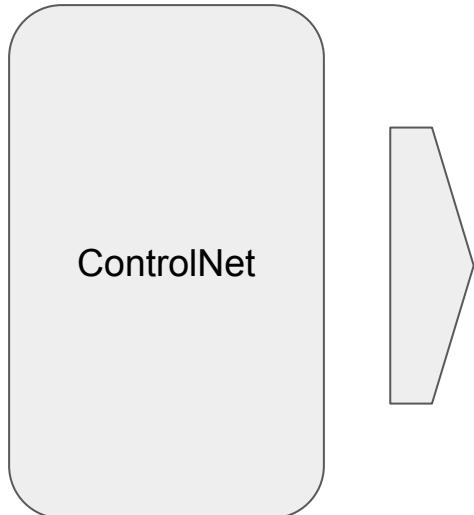
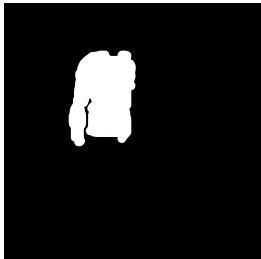


# Technical Details (Generation)

Google/Facebook Post



Mask



Prompt



# Sample Recording: peripheral view

Walk up to a crowd. Face away from the person of interest, but look at them and say,

“I like your style”

## Extracted clothing description:

Color: Green, Fabric: It appears to be a sweater fabric, likely a knit but the exact fabric content cannot be determined from the image, Length: Long-sleeved, Pattern: Graphic pattern (with what seems to be a depiction of a character or similar graphic on the front).



# Maximise Aria and Timeline Capabilities

**This is how you would look like!**

Sun May 14 2023 12:57:14 GMT+0800 (Singapore Standard Time)



[More details](#)

# Generation output



# Enhanced Input

Clothing description:

A 18 year old California cool kid with a baseball bomber jacket with cool stickers 8k highly detailed



# Inpainting + Refiner

Not So Good



# Strength 0.9

Clothing description:

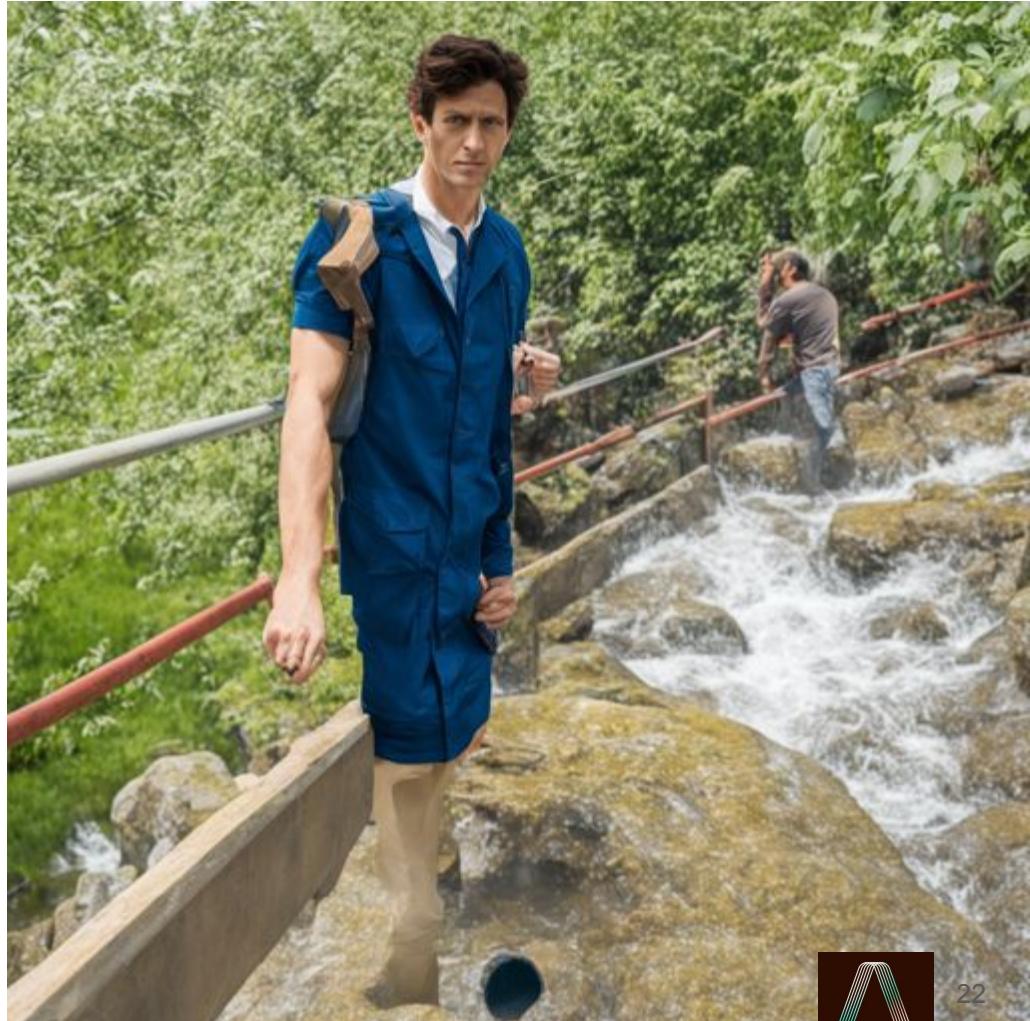
A person wearing a blue jacket



# Strength 1

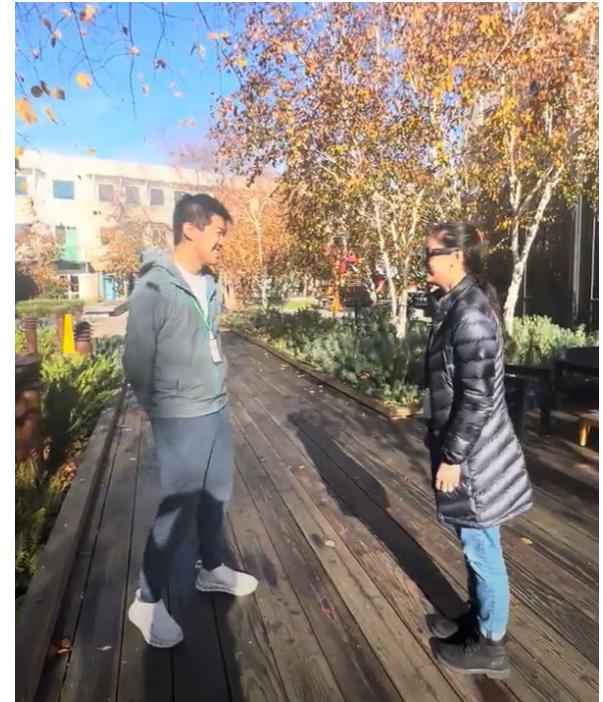
Clothing description:

A person wearing a blue jacket



# Maximize Aria Glass Tech: Features we used

- Cameras
- Eye-tracking
- Audio recorder
- Narration model + GPT4v (replaced Object Detection Model)



# Developed full end-to-end experience

## I Like Your Style

I Scanned, Magic Happened

I See It, I Want It



## I Like Your Style

I Scanned, Magic Happened

I See It, I Want It



# Developed full end-to-end experience

<https://meta-aria-front.vercel.app/>



# Ethical Model-related & Usage Considerations

- User Consent to Control (User Journey)
- Robust Authentication System to Ensure User Data Protection (User Journey)
- Data Privacy & Storage (User Journey & Monetization)
- Branded Content (AI Model Output & Monetization)
- Demographics Bias (AI Model Output & Monetization)



# What's Next

- Fuzzy search for the keyword
- Streaming
- Better controlling with control net
- Collaborative filtering between amazon data of two timelines
- Location Preferences: Personalize with commute/favourite routes

# Appendix

# Resources

Github repo

<https://github.com/vtennero/meta-aria-front>



# Why



I like your style!!

Where can I buy similar fit *for me*

Oh, I have **Aria** glasses.  
Can it help me??



# Data Acquisition

## Recordings

- Made 5 recordings with Aria glasses, where the wearer approaches people and says “I like your style”
  - Approach one or many people
  - Longer recordings
  - Person is not centered in glasses frame

## Shopping History

- Obtained from Amazon.
- Due to lack of data for ourselves, we simulated purchases based on certain personas.

# Technical Details: Backend

## Recordings:

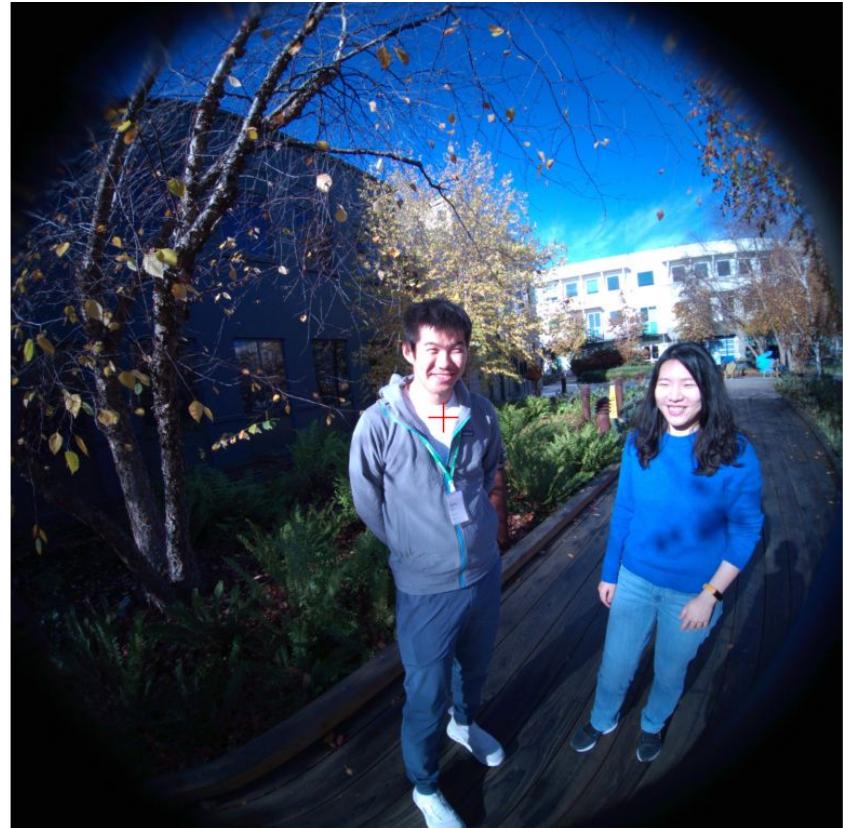
1. Take recordings of people, with the phrase “I like your style”
2. Extract speech, run speech-to-text (with timestamps)
3. Find snapshots from recordings with those timestamps
4. If there are multiple people in the snapshot at a given timestamp, run eye tracking to see who the user is focusing on
5. Run GPT4 on that person and ask it to describe what they are wearing
  - a. Prompt: “Focus on the person in the middle of the frame (selected by a red cross). Can you provide a description of the color, fabric, length, pattern of the clothes in the video? Reply in the format of Color:Black, Fabric: Cotton, Length: short, pattern: lattice”

# Sample Recording: two people

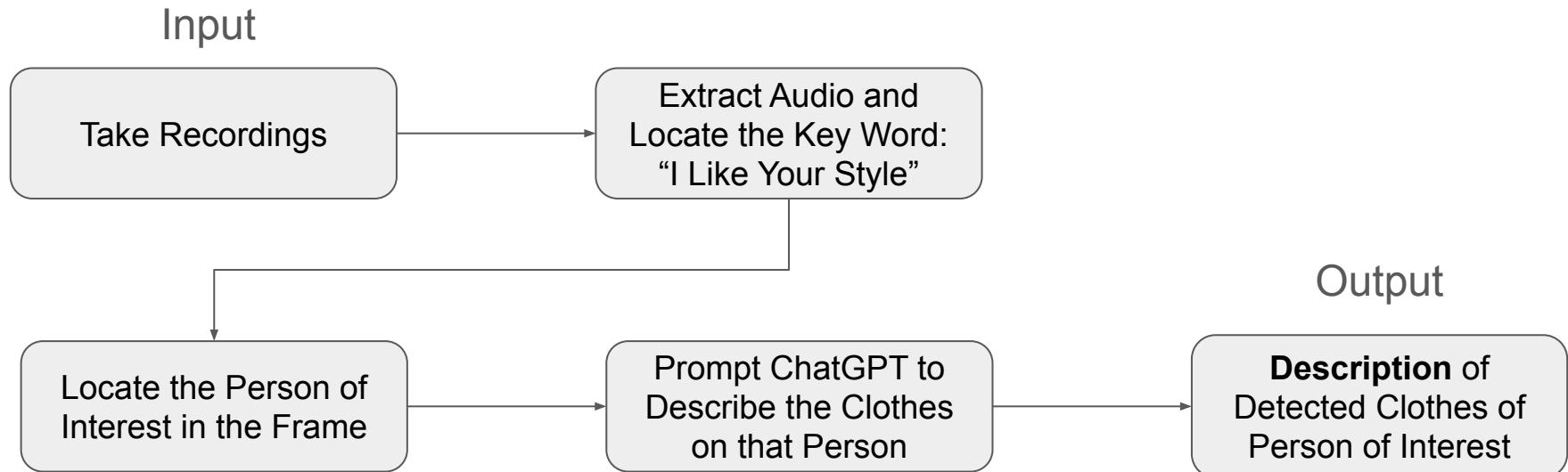
Walk up to 2 people and say to one person,  
“I like your style”

## Extracted clothing description:

Color: Blue/grey, Fabric: Synthetic blend,  
Length: Long-sleeve, Pattern: Solid



# Technical Details: Detect and Describe

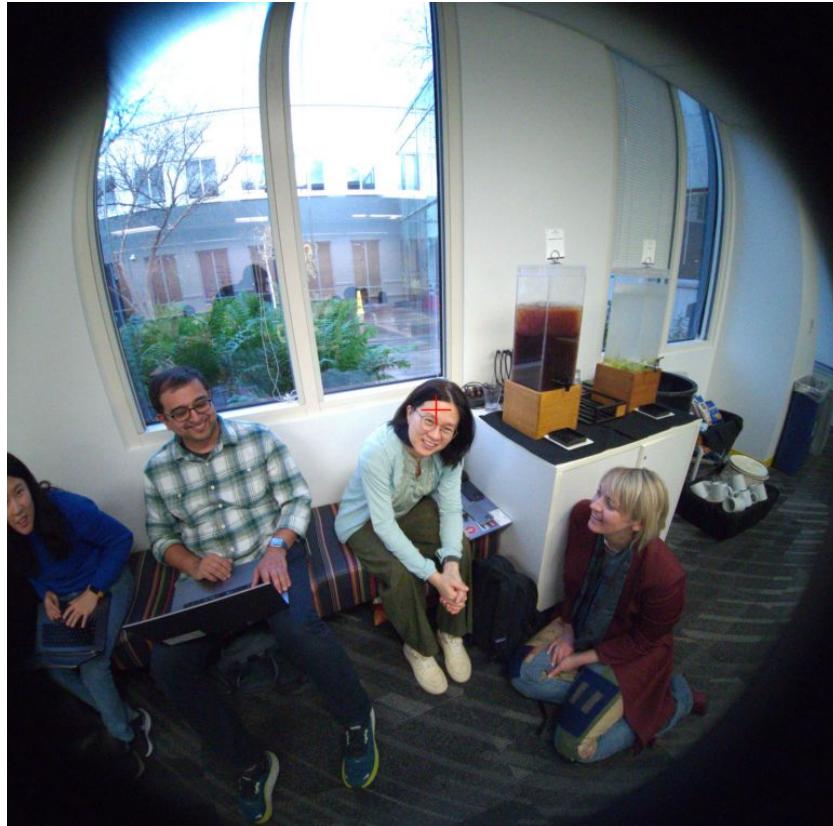


## Sample Recording 2: crowd

Walk up to a crowd and say, to one person,  
“I like your style”

### Extracted clothing description:

Color: Light Blue, Fabric: Knit (appears to be), Length: Long sleeves, Pattern: Solid



# Detected Clothing: GPT Prompt

- GPT4-v Model

"Focus on the person in the middle of the frame (selected by a red cross). Can you provide a description of the color, fabric, length, pattern of the clothes in the video? Reply in the format of Color:Black, Fabric: Cotton, Length: short, pattern: lattice"

# Findings

- Combining speech detection and eye tracking leads to high accuracy when locating person of interest
- Eye tracking helps GPT focus on the person of interest and only generate their clothing description

# Conclusions and Main Findings

-



# More Case Studies

# Recording 1: base case

Directly walk up to one person and say,  
“I like your style”

## Extracted clothing description:

Color: Grey, Fabric: Appears to be a synthetic blend, Length: Long-sleeve, Pattern: Solid with some color blocking (green and light blue details)



# Generation output



# Generation output



# Recording 5: longer instance

Walk around for 5 minutes. Go up to 2 people and, to each, say:

“I like your style”

## Extracted clothing description:

(1) Color: Black, Fabric: Appears to be cotton or a cotton blend, Length: Short-sleeved, Pattern: Graphic print

(2) Color: Black and Brown, Fabric: Leather and Fleece, Length: Waist, Pattern: Solid



# Generation output



# Generation output

