

# Visual **AI**ze: Accessible Spatial Awareness using **AI**

## Section 8

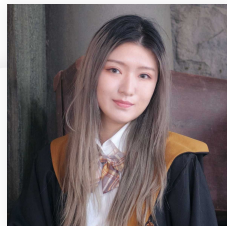
Bailey Kuehl, Violet Li, Sophia Chen, Kate Kostelc, Sacheth Swaminathan, Alyssa Nicholas

Berkeley

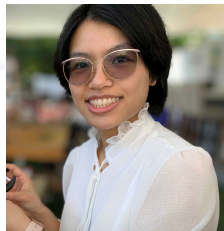
# Our Team



**Bailey Kuehl**



**Violet Li**



**Sophia Chen**



**Kate Kostelc**



**Sacheth  
Swaminathan**



**Alyssa  
Nicholas**

# Motivation

# 4 million Americans have some form of vision impairment...



Security



Signs, crowds, obstacles



Gate

**Airport travel is among the most difficult experiences for many with low vision.**

# Current airport travel process as a low vision person is "stressful" and "frustrating"

@jase123111  
I am a blind guy and I HATE airports. I have never had the confidence to use an airport by myself. When my partner or family I get so stressed in the airport just too many obstacles, noises, people and

stayinginpeace Oh trying to find your gate is such a pain! 🙄

@MostlyBlindMitch 4  
Great video. I'm a bit ashamed to admit that I missed out on an opportunity to travel to Florida for the Daytona 24hr race this year, mostly out of fear of traversing the airport and security on my own

@smmsjw  
I agree Sam... We have to be our own advocate when traveling alone.

@jamesrath  
there are multiple parties involved with providing assistance at the airport, and you never an accurate timeframe of when you'll get help. This can cause a lot of anxiety, as I was waiting for nearly 30 minutes without a sight guide showing up

@darquequeen2323  
Traveling by plane is a deep source of stress for me and I'm moving forward to it like I'd want a root canal. And trying to get someone to help can be so difficult! I'm in tears by the

# VisualAIze's Mission:

**Empower and enable people with low vision to travel independently and confidently**



**Inclusivity**



**Empowerment**



**Independence**

# Target Market is primarily individuals with low vision and smartphone access

Priority	Target Users	Rationale	Term
<b>1</b>	<b>Low vision with smartphones</b> (partial / legal blindness, retinitis pigmentosa, etc.)	<b>Easily accessible for user studies and simplest for initial development and testing. Large market with significant demand</b>	<b>Initial</b>
2	Low vision without smartphone access	Future editions can partner with accessibility services to broaden reach; currently limited by lack of smartphone access.	Long Term
3	Complete Blindness	Higher risk classification; need to prove algorithm effectiveness first;	Long Term

# Our Product

Berkeley

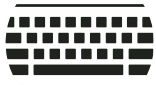


# VisualAlze's Product: Phone Application



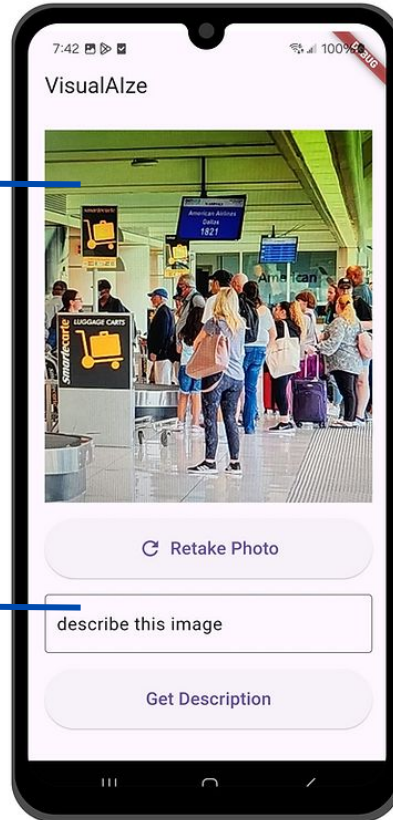
## Upload an Image

User can click this button to upload an image

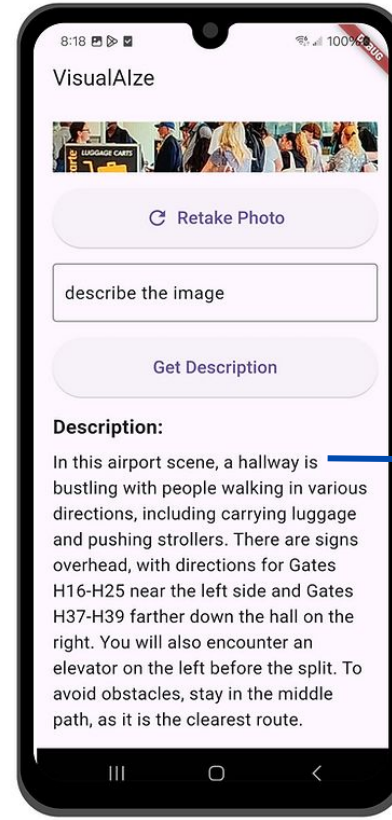


## Add a Prompt

User can click here and type to insert their own prompt



Input



## Generated Description

User can click the generated description to hear it

Output

# Key features of our product were motivated by accessibility and usability



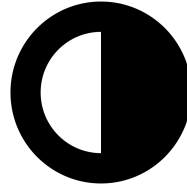
## **Audio Input / Output**

Accessibility  
Improvement



## **Increased Font Size**

Accessibility  
Improvement



## **Increased Contrast**

Accessibility  
Improvement



## **Accessibility Integration**

Accessibility  
Improvement



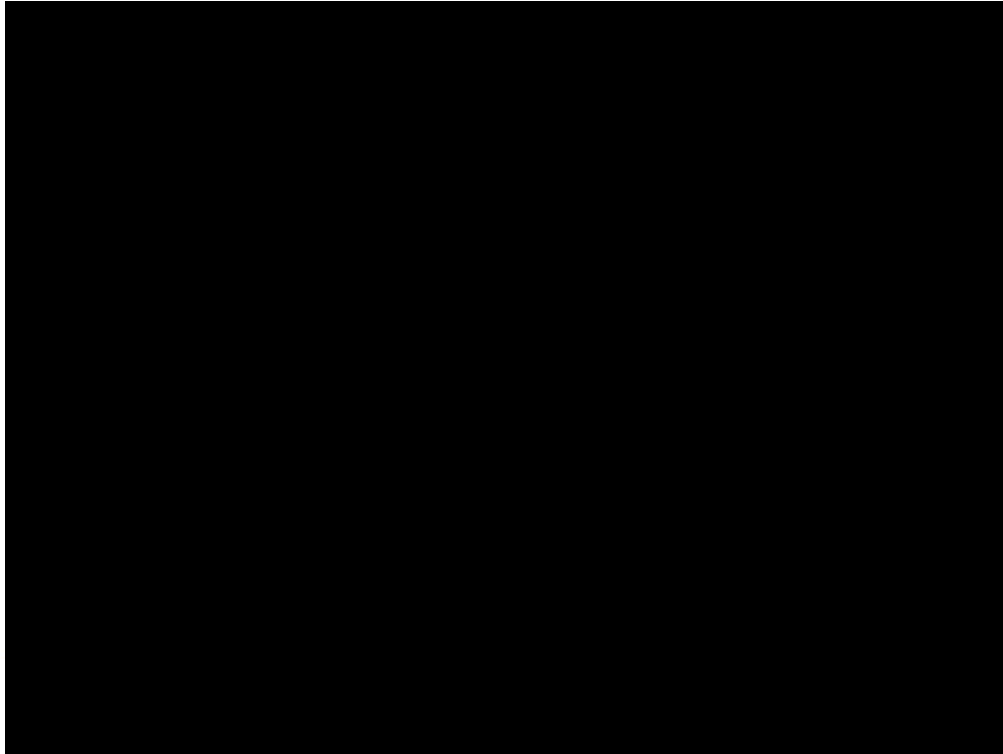
## **Consent Form**

Privacy / Legal  
Improvement

# Demo

Berkeley

# Demo: Using VisualAIze to find the departures/arrivals screen at the airport



# Data

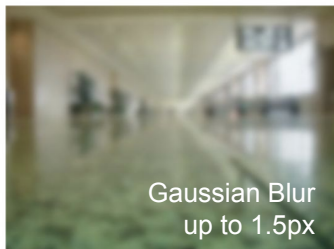
# Technical Approach - Data Acquisition and Exploratory Analysis

- **Data Source:** Indoor Scene Recognition Dataset from MIT
  - Airport indoor images
  - Image quality is unrealistic for visually impaired use case
- **EDA Performed:** Image sizes, brightness, t-SNE, and PCAs
- **Gold Data Annotation:**
  - Used 92 images (10%) to generate our golden dataset for evaluation
  - First generated by GPT, then manually reviewed / edited
  - Add highlight to **accessibility features** like tactile guide strip

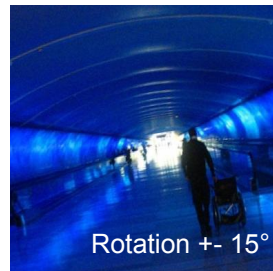


# Creating Robust and Proprietary Dataset

Applied random blurring,  
exposure.



Used Roboflow to resize and  
apply rotation.

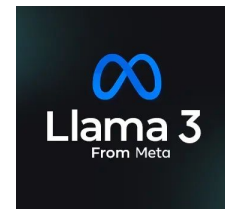


- All images stretched to 500x500 pixels

Pass images to LLM  
model like GPT4o



GPT-4o

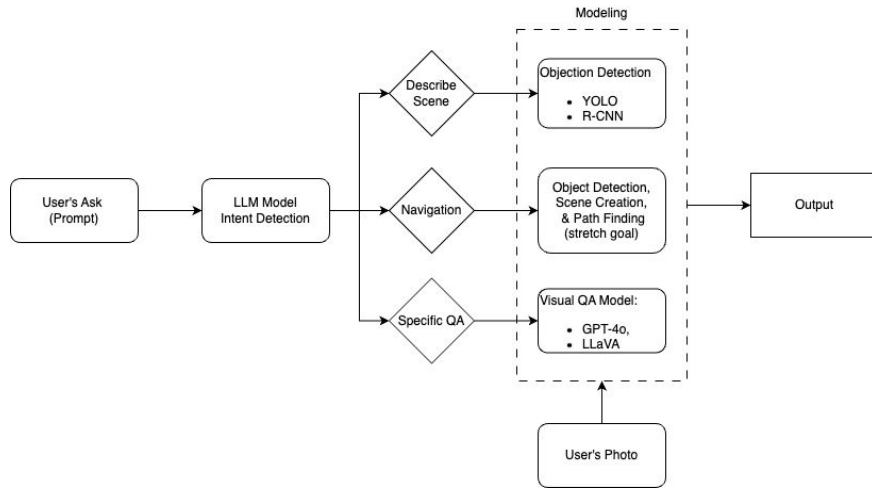


# Modeling Approaches



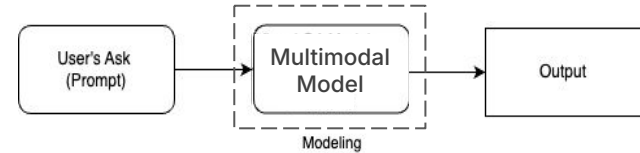
# We explored two different pipelines for our final product

## Multi-Model Pipeline

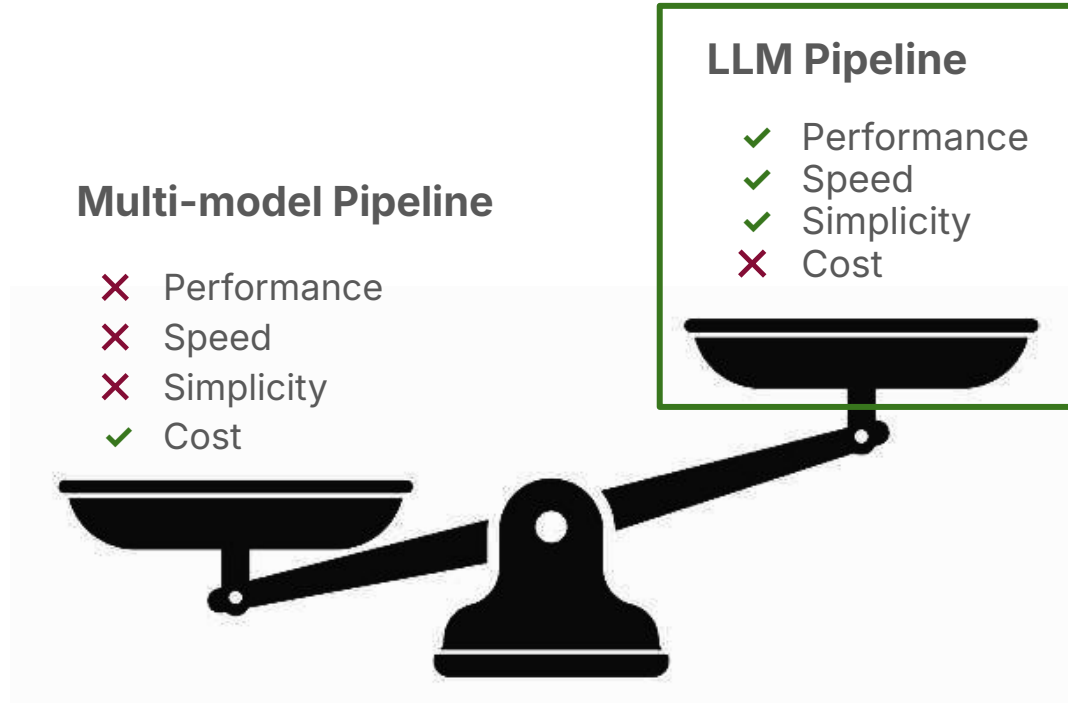


VS

## LLM Pipeline



# Using a single LLM as an end-to-end pipeline is more **efficient**, more **expensive**



# Evaluation

# Quantitative Model Evaluation: Automated Metrics

	ROUGE-L	BERTScore	Avg Runtime (sec/image)
BLIP	0.147	0.656	11.82
Llama3	0.170	0.851	13.69
gpt-4o-mini	0.187	0.868	3.16
<b>gpt-4o</b>	<b>0.189</b>	<b>0.868</b>	<b>3.31</b>

**TLDR:** gpt-4o is marginally better and faster than gpt-4o-mini on our test set for traditional metrics

# Quantitative Model Evaluation: Human Evaluation






	Human Judgements		
	<u>Concise</u>	<u>Inform</u>	<u>Accur</u>
BLIP	<b>1.00</b>	0.567	0.600
Llama3	0.00	<b>0.900</b>	0.517
gpt-4o-mini	-	-	-
<b>gpt-4o</b>	<b>0.567</b>	<b>0.878</b>	<b>0.844</b>

**TLDR:** gpt-4o did not follow instructions well on conciseness (2-3 sentences) but otherwise outperformed other models





Rubric for Human Annotators

	0 points	1 point	2 points
Conciseness	Not present or > 3 sentences.	< 3 sentences.	X
Information	More than 1 overly complex, unclear, or complicated sentences that lead to confusion. Contains complex, unclear, or unfamiliar terminology.	Lacking information to help users safely navigate their space. At least 1 unclear, complex sentence with uncommon vocabulary.	Clear and informative sentences that lead to users safely transferring spaces. Commonly used and clear vocabulary.
Accuracy	More than 1 object are incorrectly identified or hallucinated. 1 in 2 accessibility supports not identify. Failed to read signs. Or lacks information that allows user to transfers the space.	Correctly identifying objects or correctly reading signs and incorrectly one task. This could include object hallucination or incorrectly identifying objects or failing to read a sign. 1 object identified incorrectly or hallucinations.	Correctly identifying objects and their relative location, including accessibility supports (ex: tactile ground strip). Correctly reading signs.

# Qualitative Evaluation: Accessibility Features

	Image	Model Output
 BLIP		there is a long hallway with a <b>long metal railing</b> and a long wall
 Llama3		The image depicts an indoor corridor in an airport terminal. On the right, there is a <b>moving walkway</b> , commonly known as a travelator or people mover, [...] There are <b>signs overhead indicating gate numbers</b> , which assist passengers in navigating to their departure points. The flooring has a <b>tactile strip</b> , a raised pattern on the surface, often used to help visually impaired individuals detect changes in walking surfaces. [...] <i>(omitted 6 unnecessary sentences)</i>
 gpt-4o-mini		You are in a wide, well-lit corridor with smooth tile flooring. To your right, there is a <b>moving walkway</b> bordered by a <b>railing</b> , and to your left, a solid wall with a textured surface. Ahead, there are <b>signs above indicating gates</b> ; the nearest one appears to be labeled with numbers. A <b>tactile guide strip</b> runs along the floor, leading you toward the moving walkway, which may be helpful for navigation.
 gpt-4o		You are in a long, wide corridor with smooth tile flooring. To your left, there is a wall with a <b>handrail</b> running along it, and a <b>tactile guide strip</b> on the ground to help with navigation. Ahead, you can hear the sound of people and notice an escalator to your right, along with <b>signs indicating gates ahead</b> ; be cautious of the <b>moving walkway</b> on your left as you proceed.

# Qualitative Evaluation: Uninformative Images

	Image	Model Output
	BLIP	araffe roof of a train station with a sun shining through the roof
	Llama3	The image you are looking at depicts the interior of an airport terminal, characterized by a modern architectural design with a large glass roof that allows natural light to flood the space. The sun is visible through the glass, indicating it's daytime. To navigate safely through this area, be mindful of the high ceilings and the potential for reflections on the glass surfaces which could temporarily impair visibility. There are no visible signs or indicators in this particular view that would provide information about gates, restrooms, or other amenities. [...] If you're seeking specific locations like gates or restrooms, you would need to follow directional signage or ask airport staff for assistance.
	gpt-4o-mini	This image primarily shows the ceiling of an airport with large, curved glass panels allowing sunlight to stream through. There are no visible signs, people, or objects that would help in navigating the airport. <b>Please provide another image</b> that includes more details about the surroundings for better assistance.
	gpt-4o	This image displays the ceiling structure of a large, open space with a curved, glass-paneled roof, likely in an airport terminal. There are no visible signs, obstacles, or other objects that would help with navigation or indicate potential dangers. To provide a more useful description, <b>please provide an image</b> that includes more elements such as pathways, signs, or other relevant features in the airport.



# User Feedback



# Image Captioning Expert: Usefulness of Product

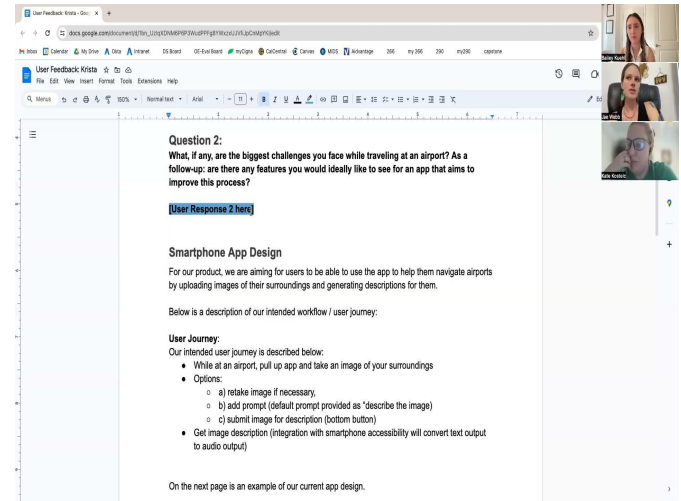
"This image description is well done. I especially like that it features information about the texture of the wall on the left, and indicates the **presence of the tactile strip.**"

- ***Cameron Bennett**, trained image alt-text writer for vision accessibility*

# Low Vision User Testimonial: Usefulness of Product

“It depends on what’s available... If I had an app, just to give me that **sense of peace, security**, knowing that it’s fine, I can use the app... I would definitely use that feature.”

- *Krista W., retinitis pigmentosa and legally blind with <math><5^\circ</math> of central vision remaining*



The screenshot shows a Google Docs document titled "User Feedback Krista" with a video call overlay on the right side. The document content includes:

**Question 2:**  
What, if any, are the biggest challenges you face while traveling at an airport? As a follow-up: are there any features you would ideally like to see for an app that aims to improve this process?

[User Response 2 here](#)

**Smartphone App Design**  
For our product, we are aiming for users to be able to use the app to help them navigate airports by uploading images of their surroundings and generating descriptions for them.

Below is a description of our intended workflow / user journey:

**User Journey:**  
Our intended user journey is described below:

- While at an airport, pull up app and take an image of your surroundings
- Options:
  - a) retake image if necessary,
  - b) add prompt (default prompt provided as "describe the image")
  - c) submit image for description (bottom button)
- Get image description (integration with smartphone accessibility will convert text output to audio output)

On the next page is an example of our current app design.

# **Learnings and Conclusions**

# Key Technical Takeaways

## **Takeaway 1: Human Involvement with AI model deployment**

- Observed that data has to be preprocessed to ensure high-quality inputs for the models.
- Model outputs has to be evaluated by human to ensure nothing is missed.

## **Takeaway 2: Compound Gen AI Model Evaluation and Deployment**

- Conducted extensive hyperparameter tuning, model evaluation, and cost analysis to confirm that GPT-4o is the model we will deploy.
- Paid for the API to save time and reduce complexity.

## **Takeaway 3: Accessibility Features Integration**

- Developed and integrated multiple accessibility features such as audio input/output, increased font size, and dynamic prompting to enhance user experience.
- Also highlighted the accessibility features within the scene.

# Future Directions

## Expand Dataset

Collect and incorporate more diverse datasets to improve model robustness and accuracy across various environments.

## Larger User Study

Take place at airport during user travels with Volunteer Users

## Incorporate Feedback

Using feedback and description evaluations to improve the model.

## Update Product

Adjust and improve product to support user experience

# VisualAIze's Mission:

**Empower and enable people with low vision to travel independently and confidently**



**Inclusivity**



**Empowerment**



**Independence**

# Appendix

Thank you Kira and Joyce and our classmates for your support, directing, and feedback. You helped to make the product better.

[VisualAlze Website](#)

Demo Video

[Ischool Website](#)

# Low Vision User Testimonial: Usefulness of Product

"If I had an app, just to give me that **sense of peace, security**, knowing that it's fine, I can use the app... I would definitely use that feature."

- ***Krista W.**, retinitis pigmentosa and legally blind with  $<5^\circ$  of central vision remaining*



# Success Rate of Key Qualitative Features on our final model (gpt-4o)

	Identifies accessibility features <i>when present</i> (True Positive %)	Requests new image on uninformative images (True Positive %)
<b>gpt-4o</b>		

**TLDR:** gpt-4o successfully identified the need for more information