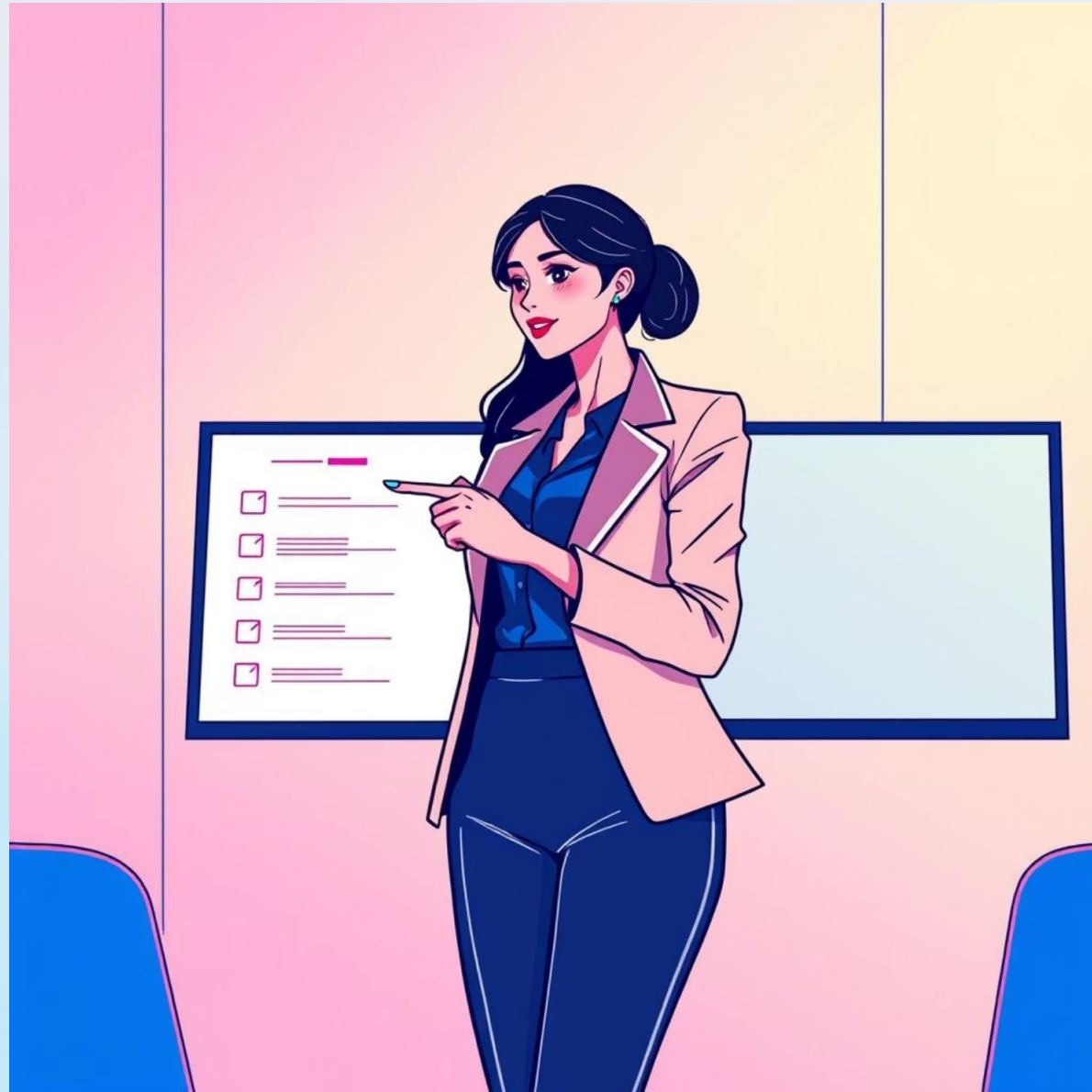


Practical Applications of GenAI

Guardify AI

Transforming Loss Prevention with Real-time AI

Agenda



- Introduction
- UI Demo
- Application Development
- Methods
- Results
- Summing Up

Motivation: The Genesis of Guardify AI

- Our journey began during a casual Friday night dinner, sparking a conversation about the escalating challenges retailers face with shrink.
- This discussion catalyzed the idea for [Guardify AI](#): a generative AI application designed for real-time shoplifting detection.
- Our goal is to provide a proactive, intelligent solution that goes beyond traditional surveillance.



Core Objectives

1

Empower Retailers

Alleviate concerns about theft by providing a robust, intelligent monitoring system.

2

Enhance Awareness

Increase real-time visibility into suspicious activities, enabling immediate response and prevention.

3

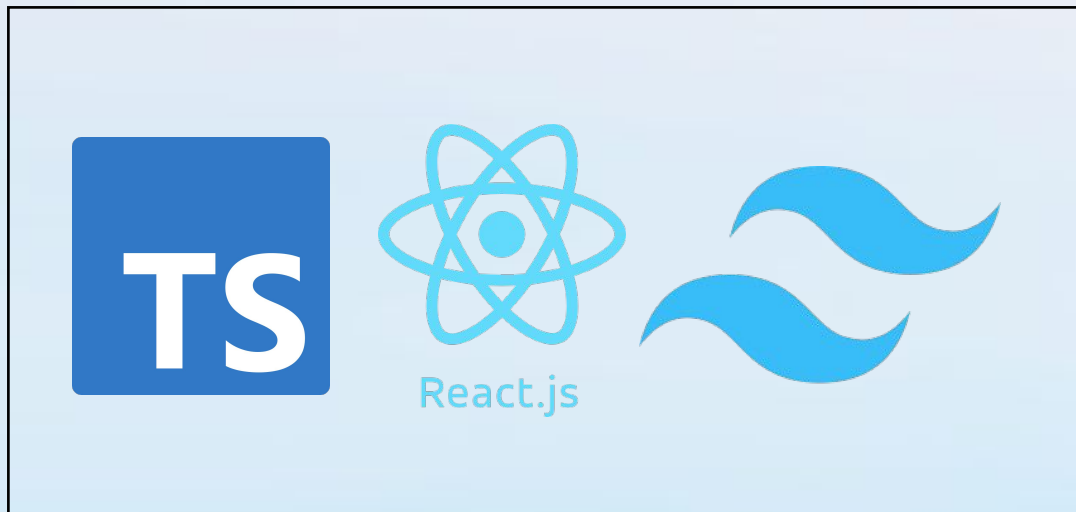
Optimize Operations

Streamline security processes and reduce the need for constant human surveillance. This allows staff to focus on customer service and other critical tasks.

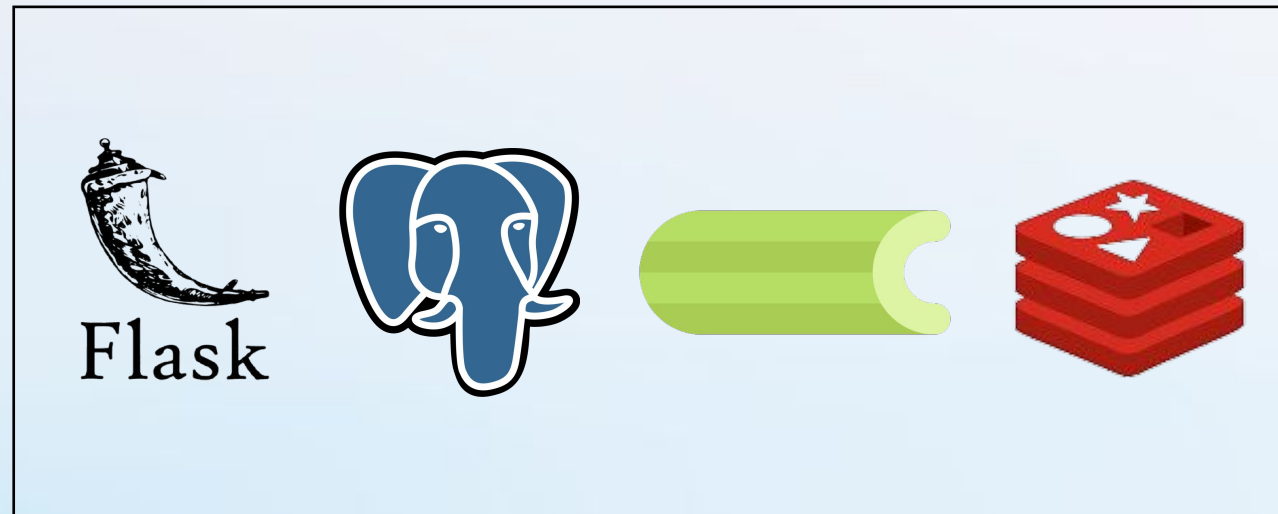


Application Flow: How Guardify AI Works

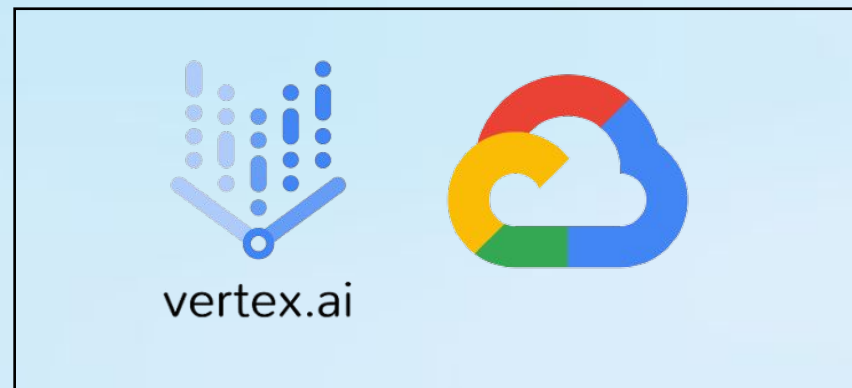
Frontend



Backend

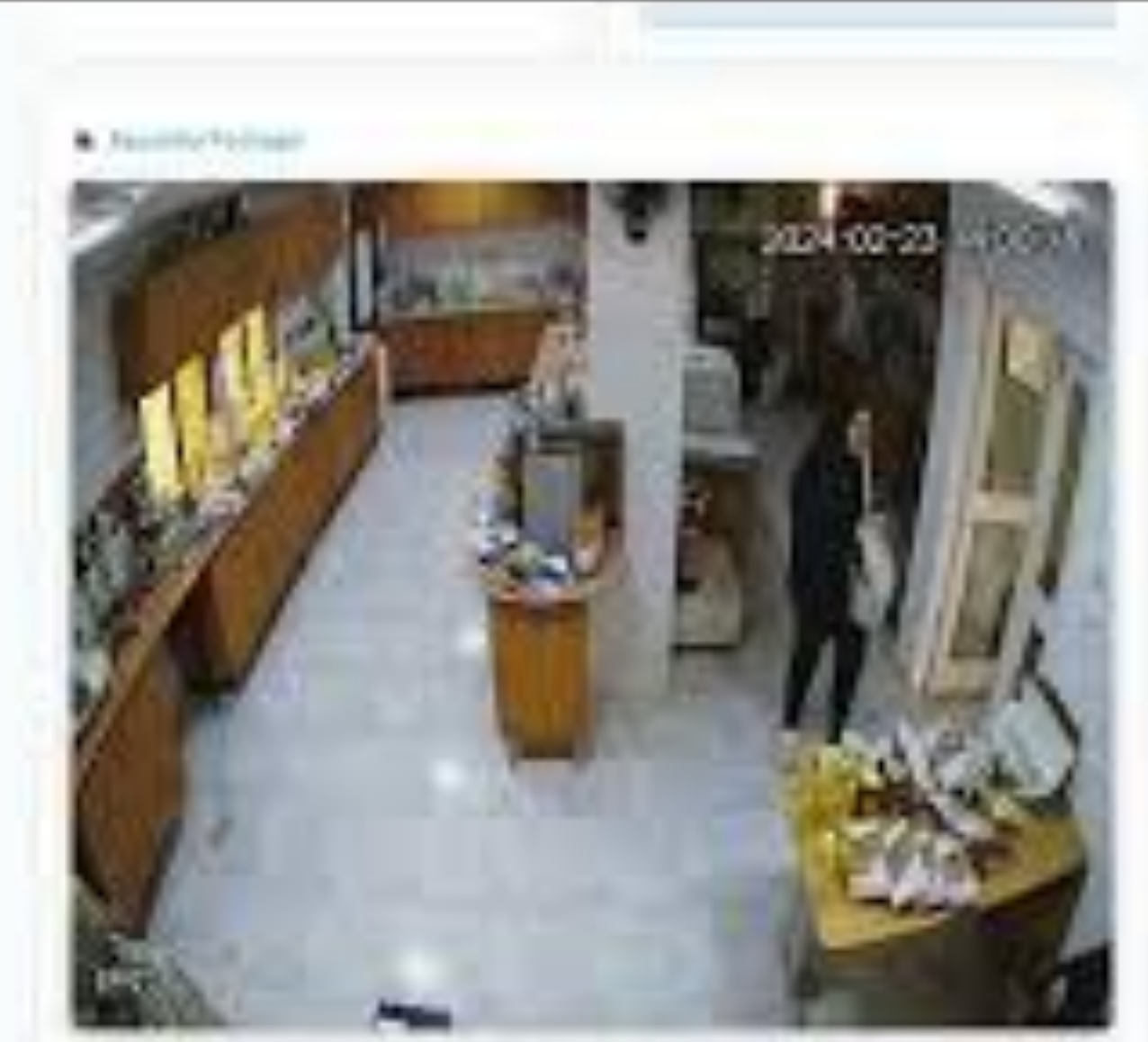


Data Science



Now demo ... (:

UI Demo: Event Page - Shoplifting Detected

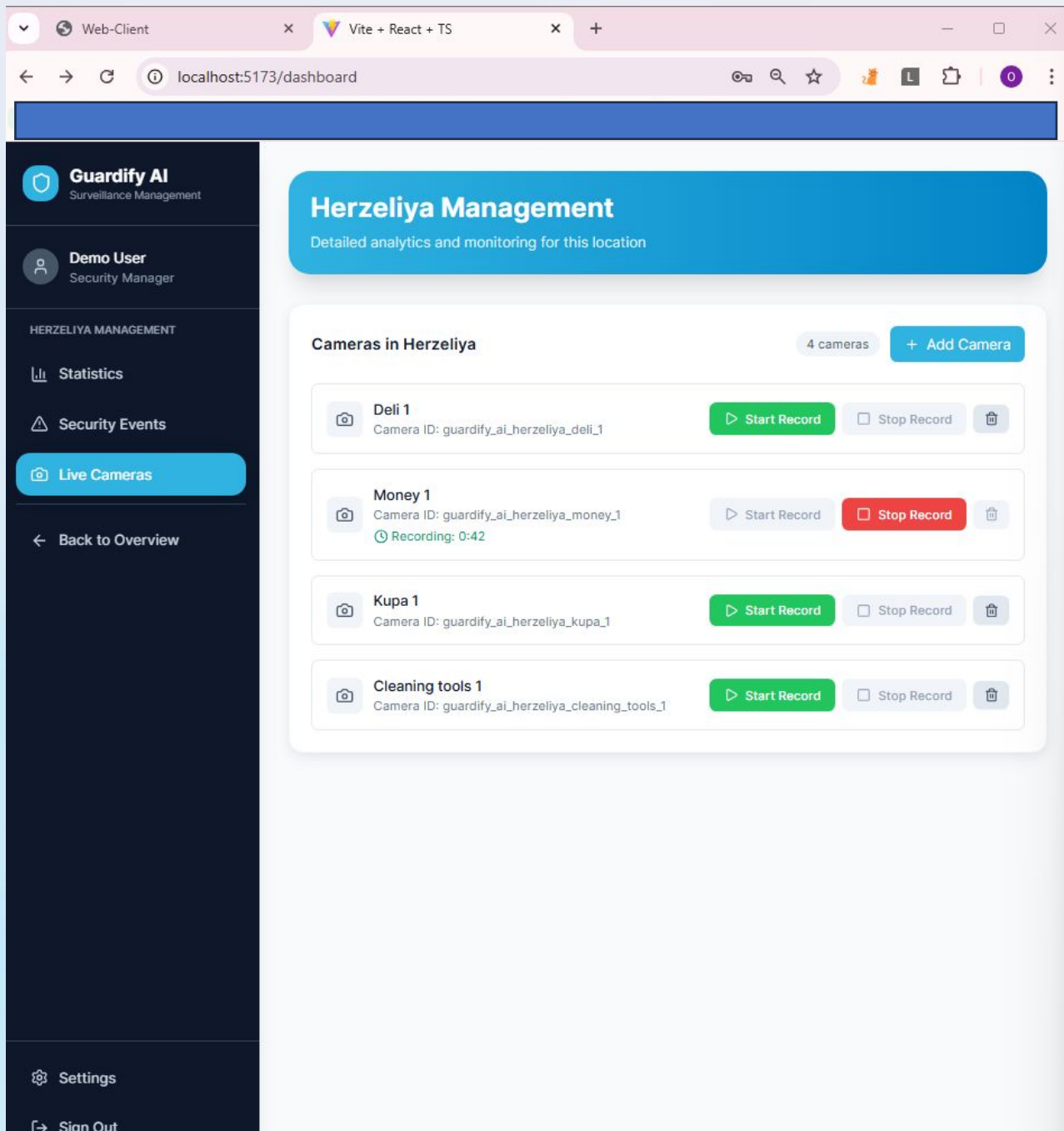


UI Demo: Event Page - Non Shoplifting

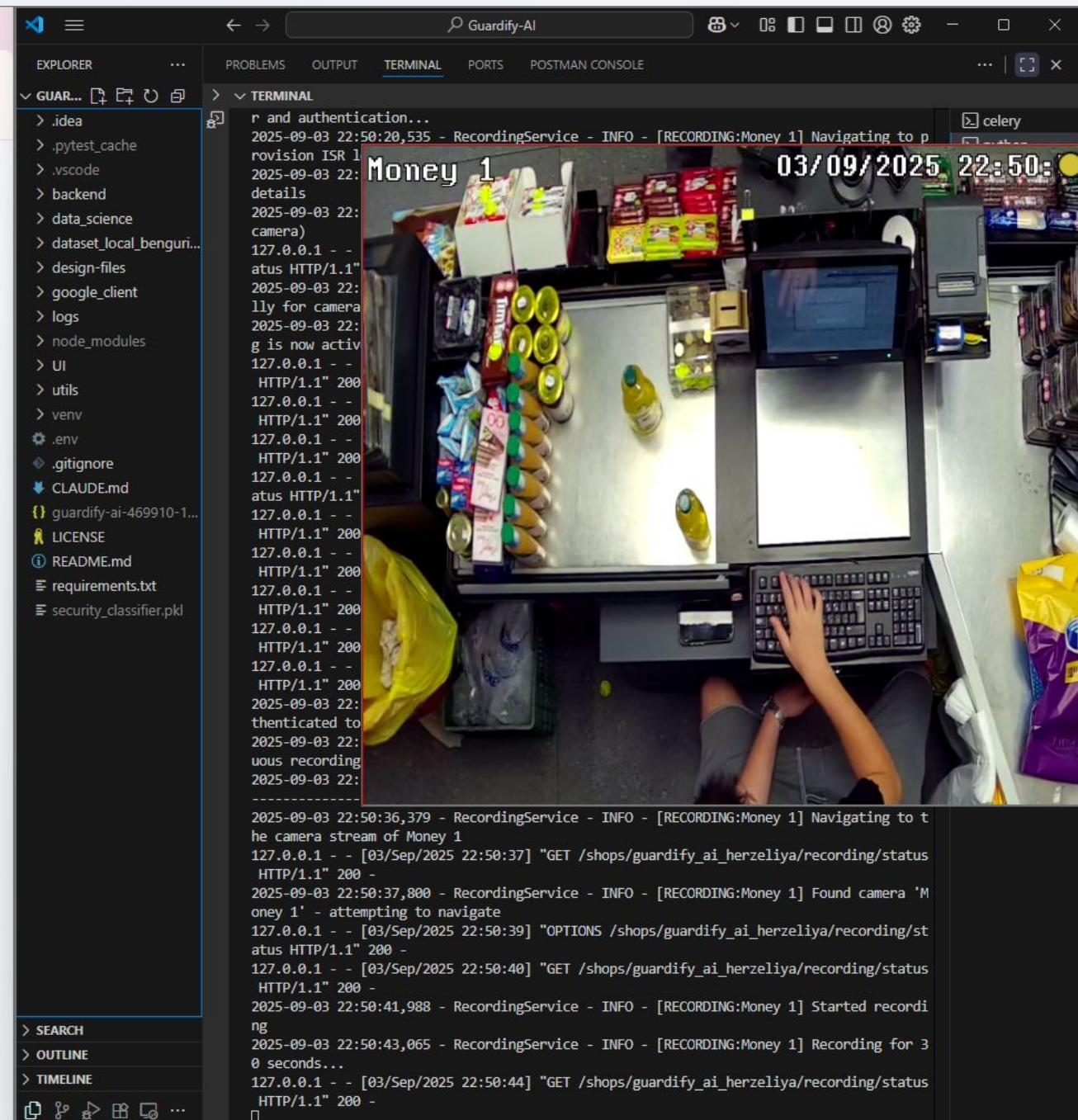


UI Demo: Real-time Incident Management

Frontend



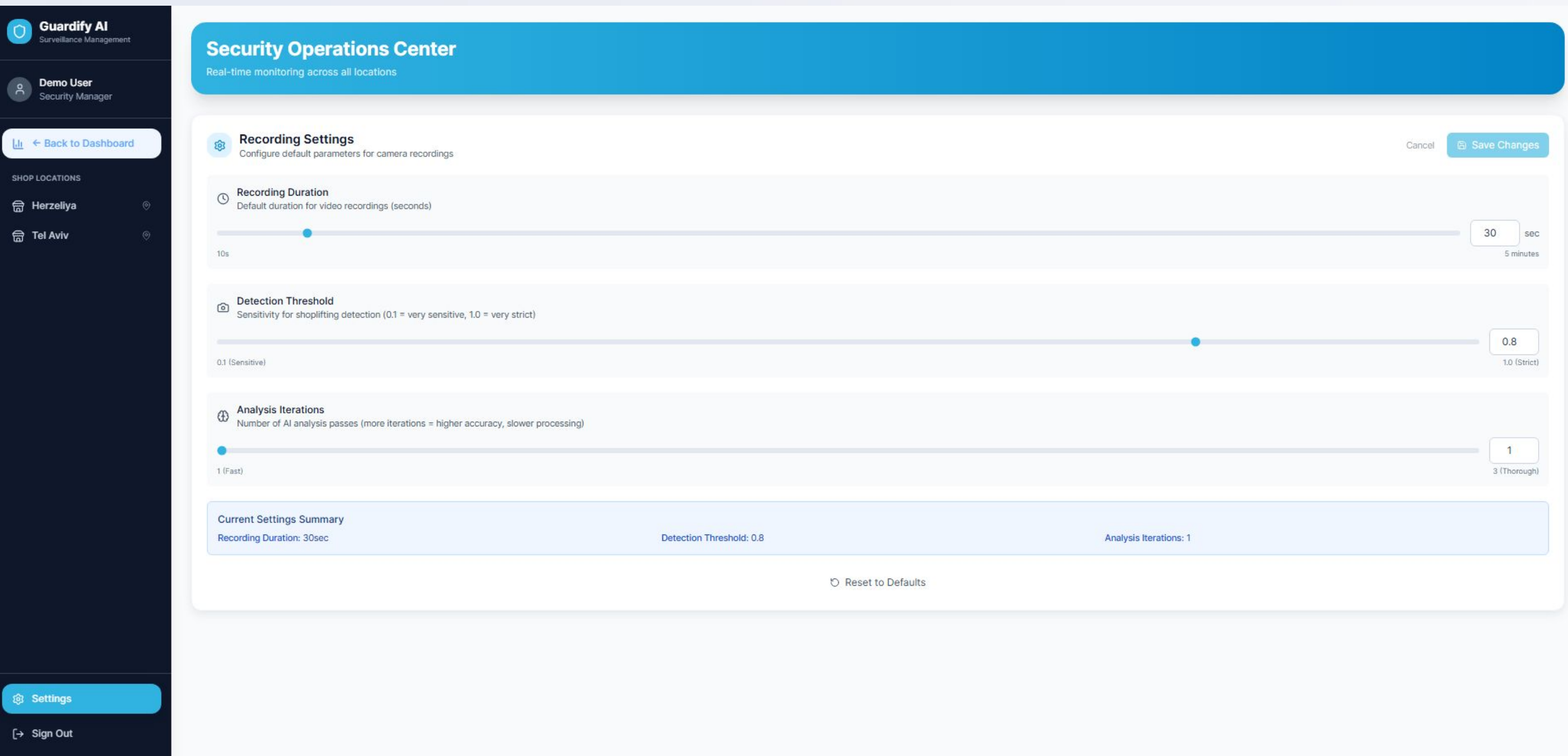
Backend



UI Demo: Real-time Incident Management



UI Demo: Real-time Incident Management



From Backend Side

Video Upload →

- Contains 2 services: Recorder & Uploader, which run in different threads for real time

Redis Queue →

- Acts as a message broker between video upload and analysis workers

Celery Analysis →

- Picks up analyze video task from Redis analysis queue
- Triggers the agentic model: Computer Vision → LLM decision making

DB Objects (Event + Analysis) →

- Generates event record with AI-powered description
- Stores detection results

UI Dashboard

- Task returns structured result with event ID and analysis data

GenAI Requirements & Model Evolution

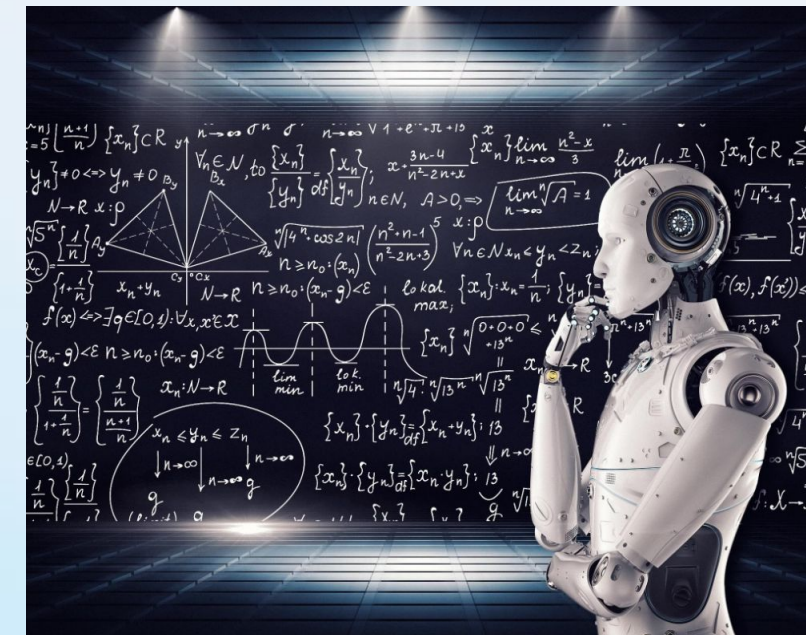
Our journey began with Azure and OpenAI models, but evolved to optimize for performance and scalability.

Initial Exploration (Azure + OpenAI):

- GPT-4o (OpenAI) for complex scene understanding.
- Phi-3 Vision (Microsoft) for object recognition.
- Whisper (Voice analysis) for audio context.

Transition to Google Cloud Platform (GCP):

- Leveraging Gemini-Flash 2.0 and 2.5 for enhanced vision capabilities.
- Exploring fine-tuned models for domain-specific accuracy.

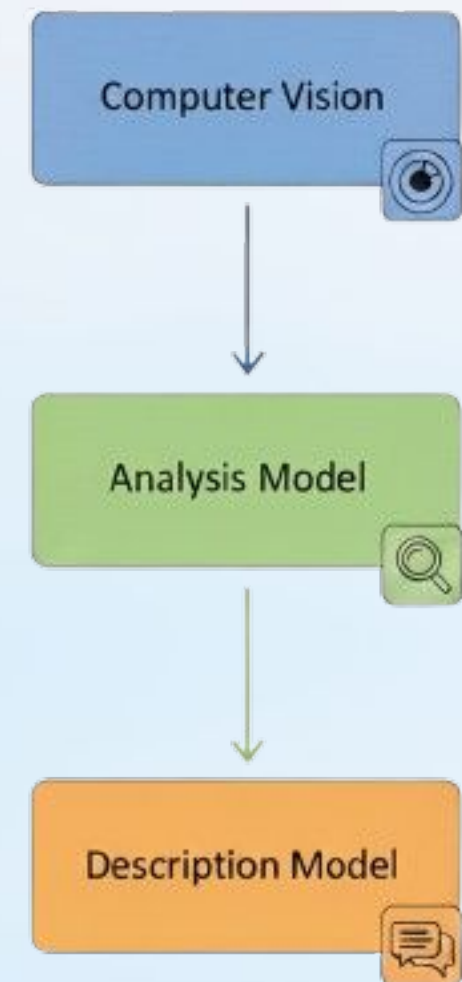


This transition allowed us to better handle the demanding real-time processing needs of Guardify AI.

Agentic Solution

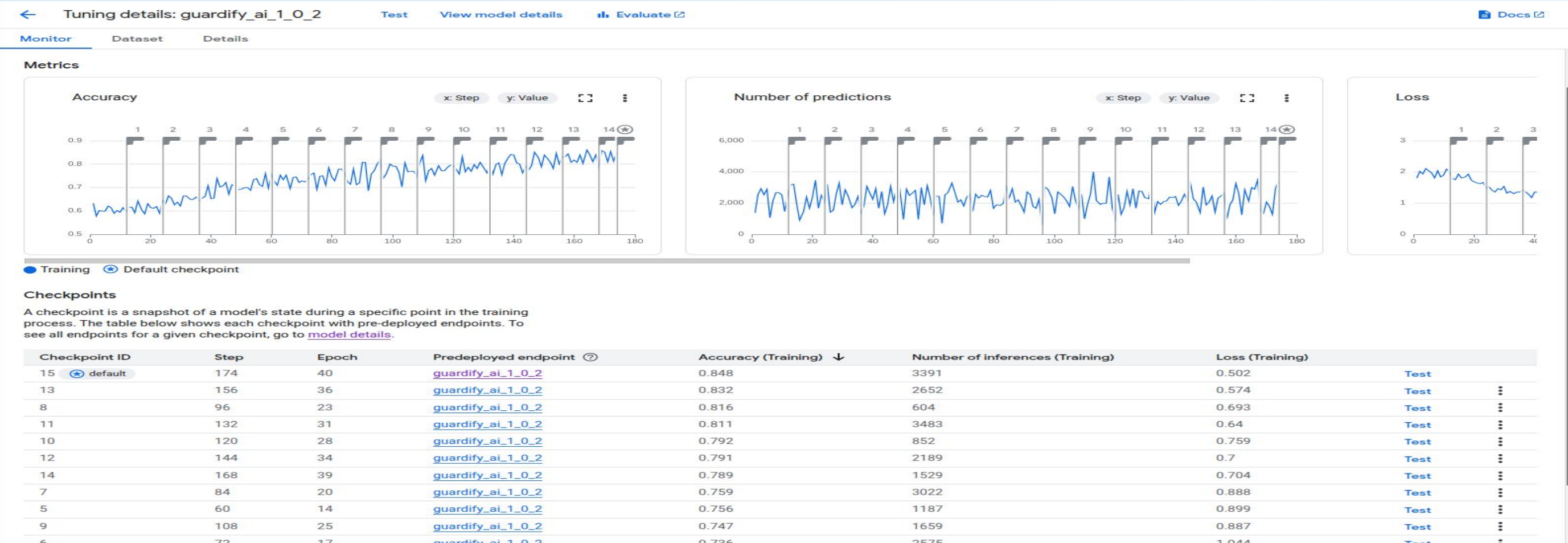
Selecting the right model architecture was critical. We developed an agentic workflow to balance speed and accuracy.

- **Computer Vision Model:** Provides detailed, structured visual observations of customer behavior, without making theft judgments.
- **Analysis Model:** Processes the structured observations from the CV model to make intelligent theft detection decisions using enhanced decision logic, confidence calibration, and protection against false positives.
- **Event Description Model:** Generates concise, professional descriptions of detected events from the analysis reasoning.
- **Advantage:** Higher accuracy due to specialized model functions.



Fine Tune Process

- Collected 52 real-world shop videos with durations aligned to our application use case.
- Fine-tuned Gemini Flash 2.5 using Google Vertex AI platform.
- Experimented with two approaches: Splitting videos into individual frames and Using full video sequences
- Conducted 8 training iterations, exploring and optimizing different hyperparameters.
- After each iteration, the model was re-tested on a validation set, with precision, accuracy, and text quality evaluated and recorded.



Model Detection Rate

TEST	Amount	Accuracy KPI	gpt-4o	Gemini Lite	Gemini	Fine Tuned Gemini (Result of validation set)
Store of Interest	53	70%	52%	50%	68%	90%
Kaggle Dataset	924	70%	46%	45%	75%	96%

SUMMING UP

- Real-time anomaly detection
- Scalable, cloud-based processing
Seamless integration with existing systems
- Our vision:
Empowering retailers to stay ahead of theft, optimize operations, and focus on what matters most – their customers.

Questions?

