AI Engineering Lab
Week 2

Haidar Chaito & Noureddine Kamzon

September 12, 2025

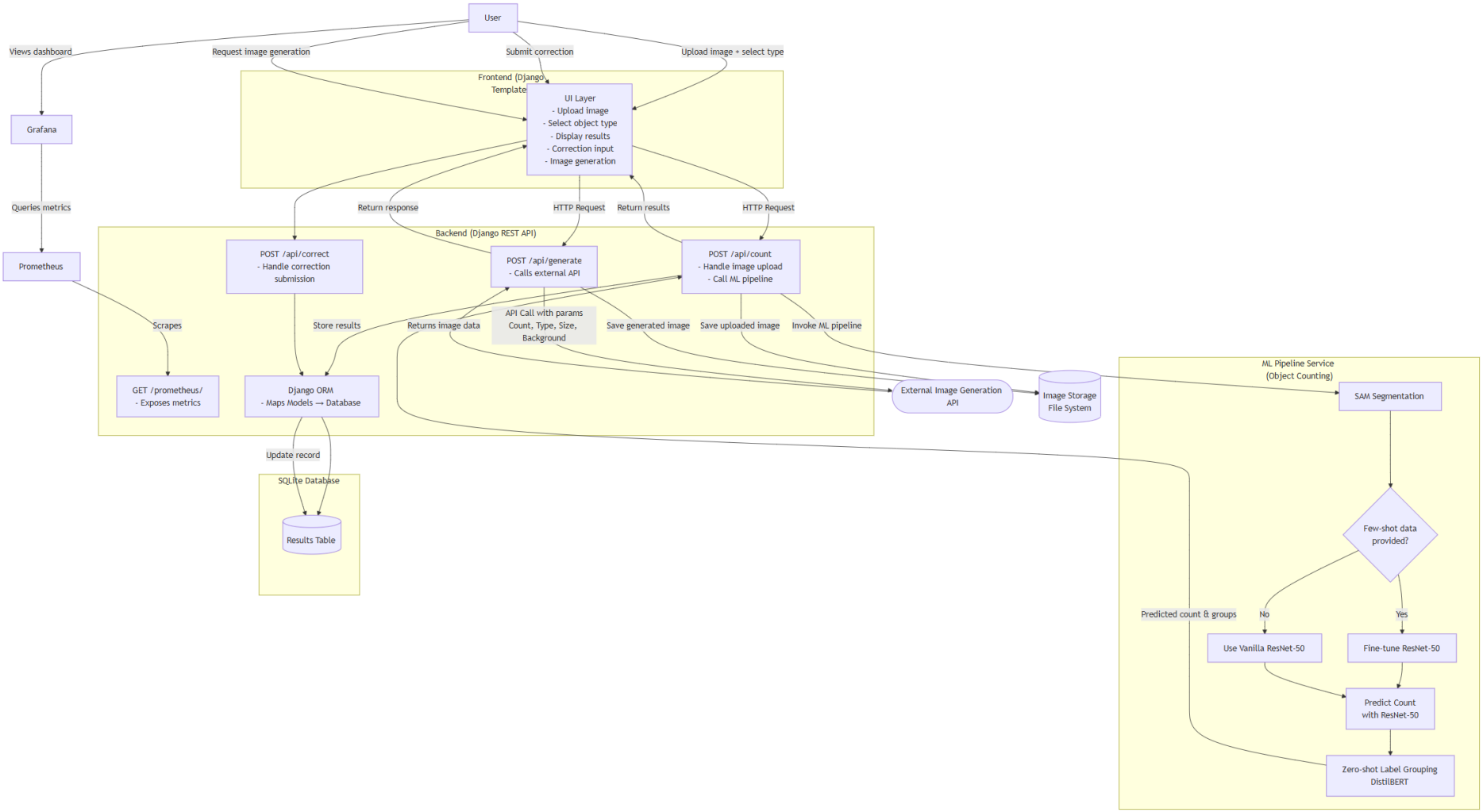
Project Goals

- Few-shot learning fine-tuning the ResNet-50 model
- Auto image generation using the API call
- Monitoring using Grafana and Prometheus
- Automated API calls

Key Concepts & Definitions

- **Few-shot Fine-tuning** Adapting a pre-trained model to a new task using a very small number of examples.
- **ResNet-50** A 50-layer deep neural network for image classification, known for its effective "residual connections".
- **SAM (Segment Anything Model)** A versatile foundation model from Meta AI that can identify and outline virtually any object in an image.
- **Distilbert model** A compact Transformer model fine-tuned for Natural Language Inference (NLI), enabling it to perform zero-shot classification by determining which label a given text logically implies.

System Architecture



Monitoring: Key Metrics

Input & Object Metrics

- Image Resolution (W/H)
- Number of Segments
- Count of Objects
- Number of Object Types
- Objects by Type (Pie Chart)
- Avg. Segment Resolution

Model Metrics

- Models Used
- Classifier Min. Confidence
- Confidence per Segment
- Accuracy
- Precision
- Recall

Performance Metrics (ms)

- Overall Response Time
- SAM Inference Time
- Classifier Inference Time
- Zero-shot Inference Time

Monitoring Implementation

- **Prometheus:** An open-source toolkit used to collect the application metrics as time-series data.
- **Metric Exposure:** The Django backend exposes its metrics using the `django-prometheus` library.
- **Endpoint Configuration:** This is configured in the Django URL patterns to serve metrics at a specific path.
- **Grafana:** The collected data is then queried from Prometheus and visualized in the Grafana dashboard.

References

- PyTorch Documentation — `https://pytorch.org/docs/stable/`
- Segment Anything (Meta AI) — `https://segment-anything.com/`
- Hugging Face Transformers — `https://huggingface.co/docs/transformers`
- Prometheus Documentation — `https://prometheus.io/docs/`
- Grafana Documentation — `https://grafana.com/docs/`