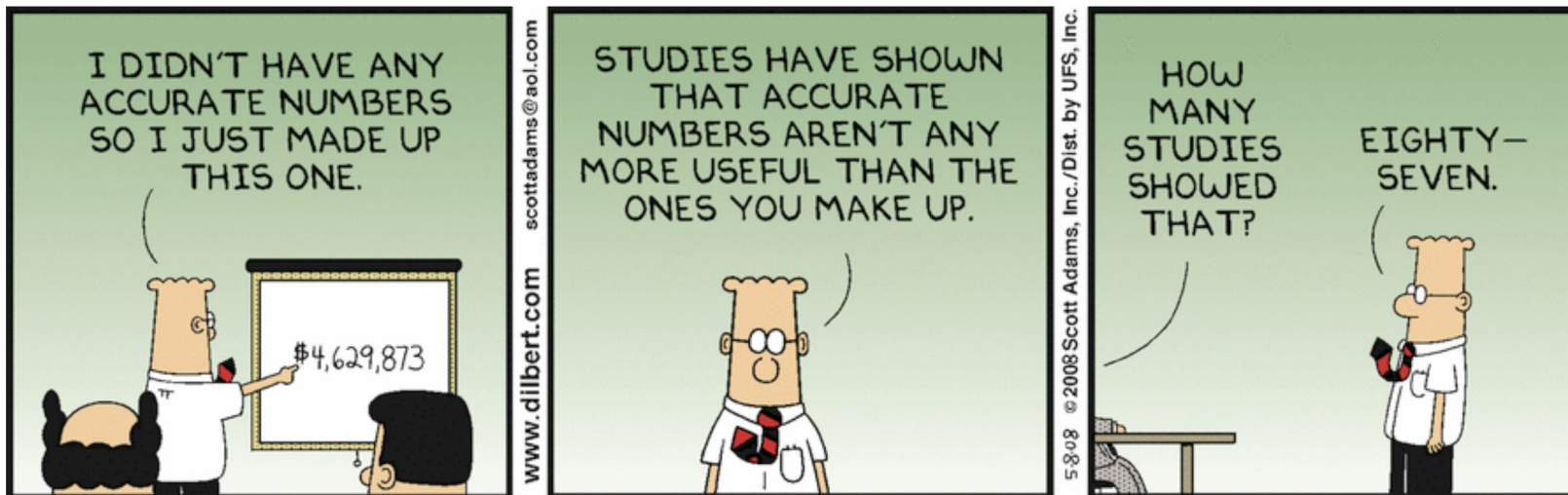


Fake News Detector

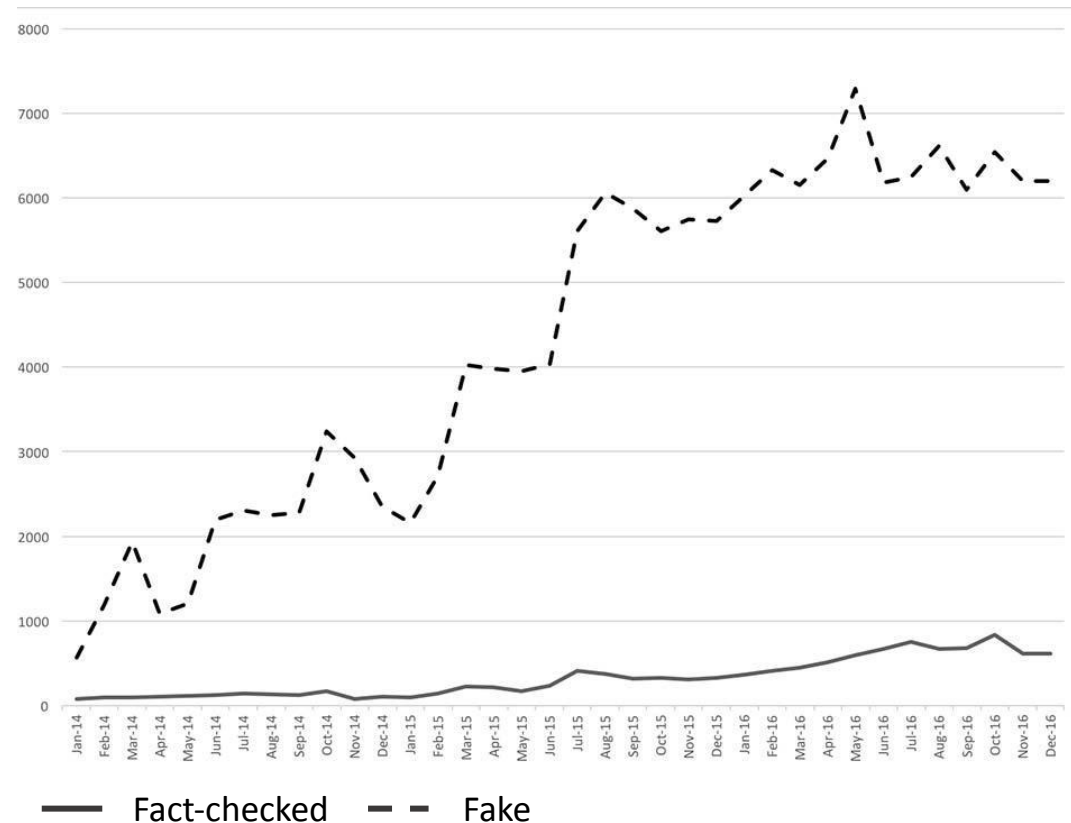
Kyle Ke, Boshen Yan, Fuchen Li, Zihan Wang, Qassi Gaba



Problem | Fake news is becoming more prominent

Prevalence of fake news is increasing

Number of stories/articles



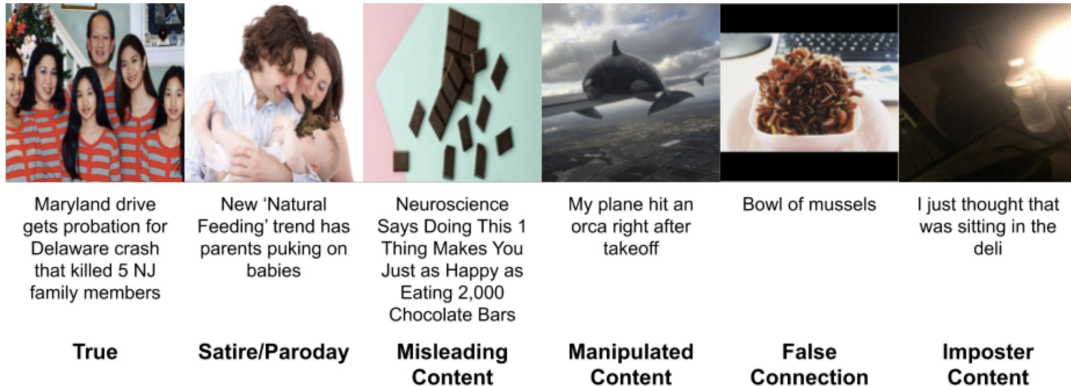
Key driving factors

1. Increased ability to create and distribute fake news
2. Inability to detect fake news

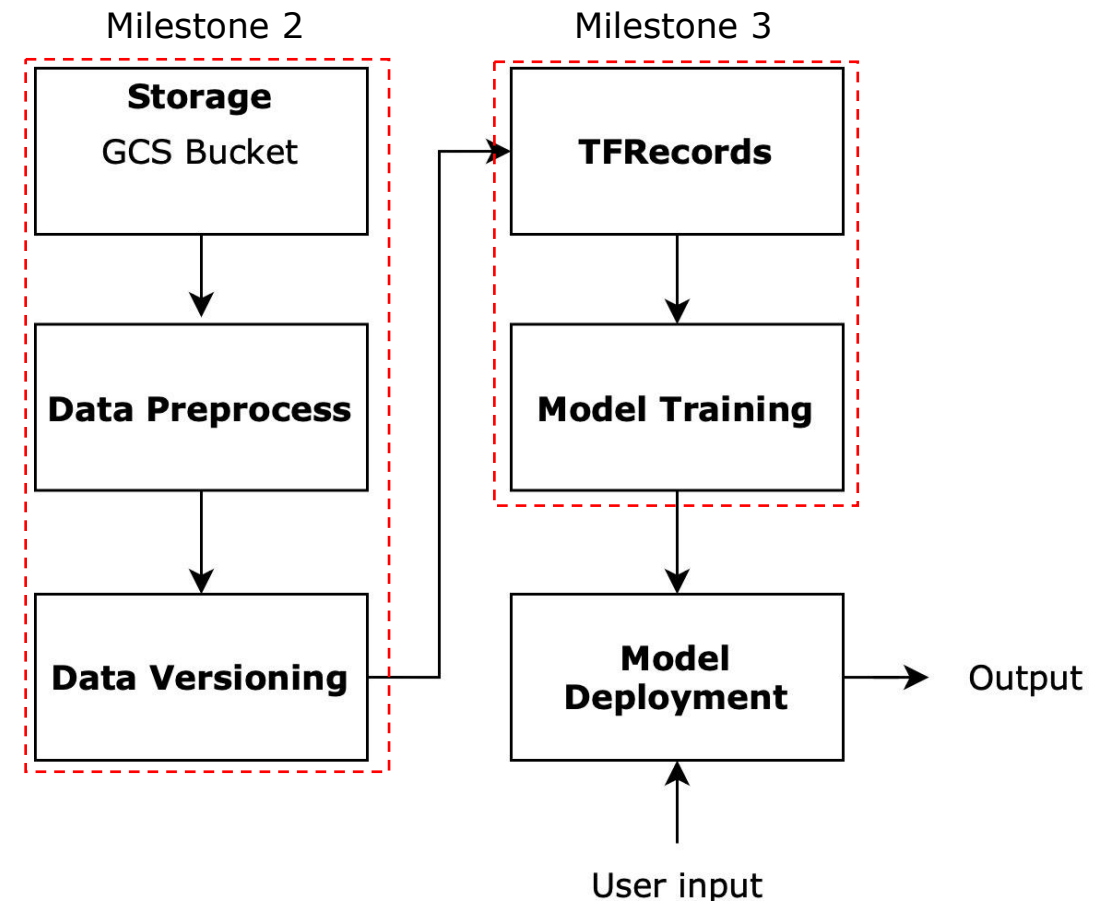
Solution | A multi-modal fake news detector

Project overview

- **Project:** Fake news detector
- **Inputs:** text + image data
- **Output:** classification in one of six categories



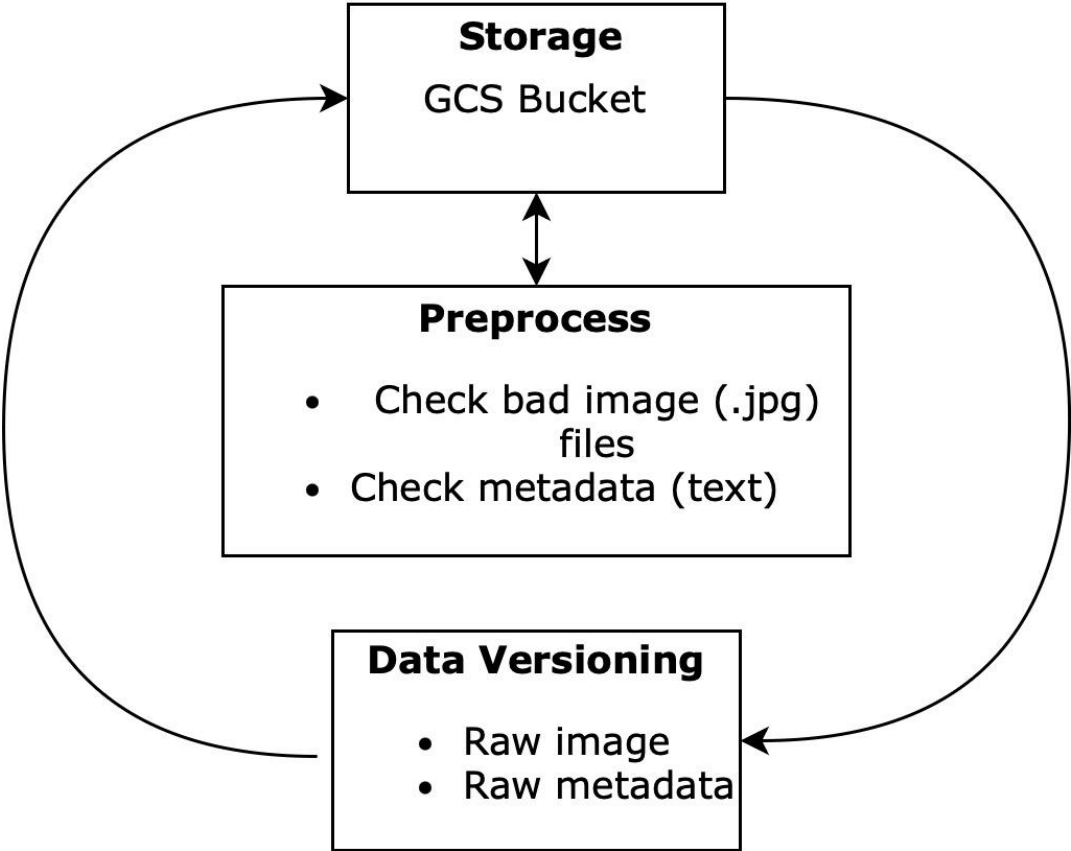
Project pipeline



* Upload and download at each step from GCP storage (not shown in diagram)

Milestone 2 | Storage, preprocessing, and versioning

Milestone overview

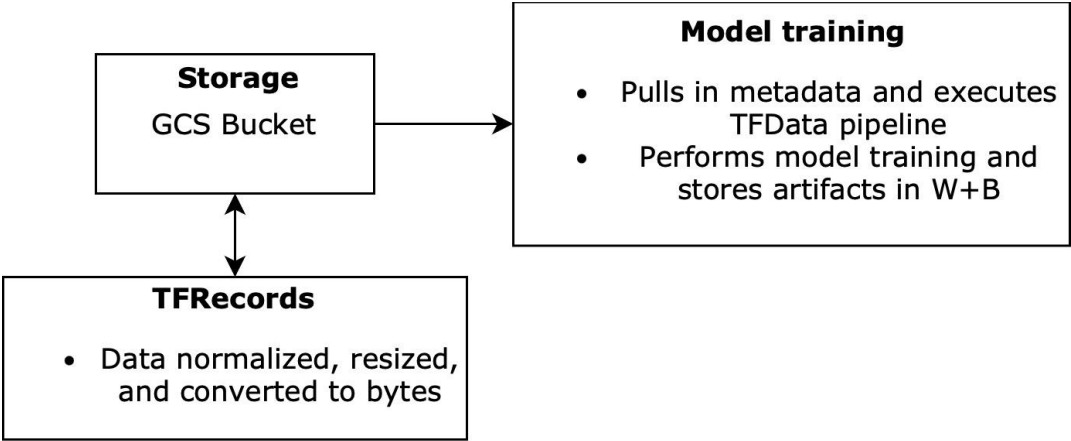


Key challenges and solutions

Challenge	Solution
Large dataset	<ul style="list-style-type: none">• Upload processed metadata only to GCP• Upload in .zip file
Poor quality of data	<ul style="list-style-type: none">• Drop NA• Remove corrupted image index from metadata

Milestone 3 | Storage, TFRecords, and model training

Milestone overview

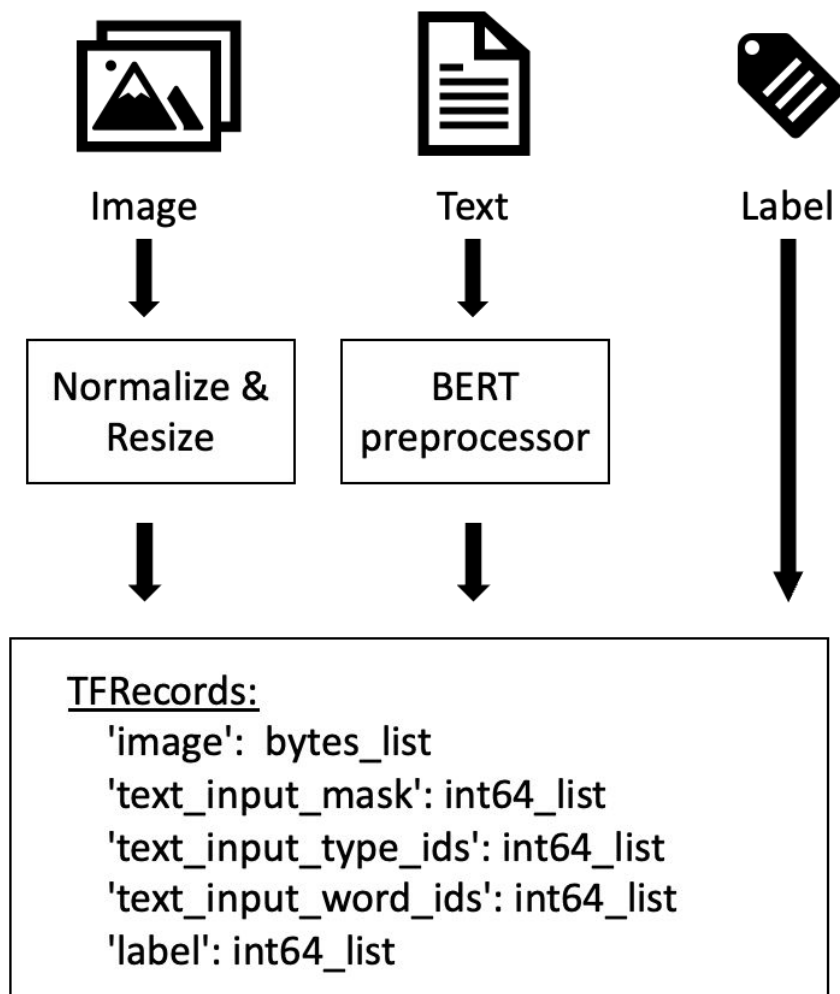


Key challenges and solutions

Challenge	Solution
Pre-defined training/test split	<ul style="list-style-type: none">Optimize model performance by adjusting hyperparameters
High GPU requirement for model training	<ul style="list-style-type: none">Leverage multi-GPU training

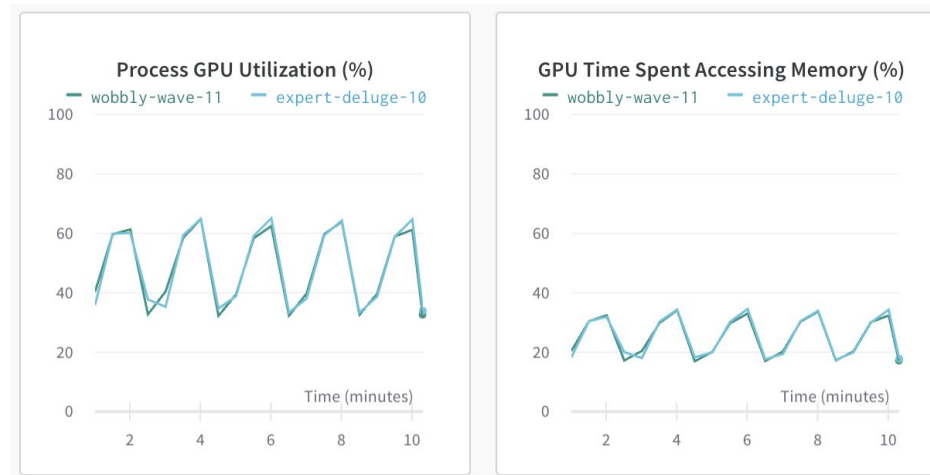
To TFRecords, or not to TFRecords, that is the question

TFRecords container overview



Issues

1. Slow to generate - 500 samples / min
2. Preliminary testing had similar speed - more testing to check if GPU utilization improves compared to pre-fetched TFData

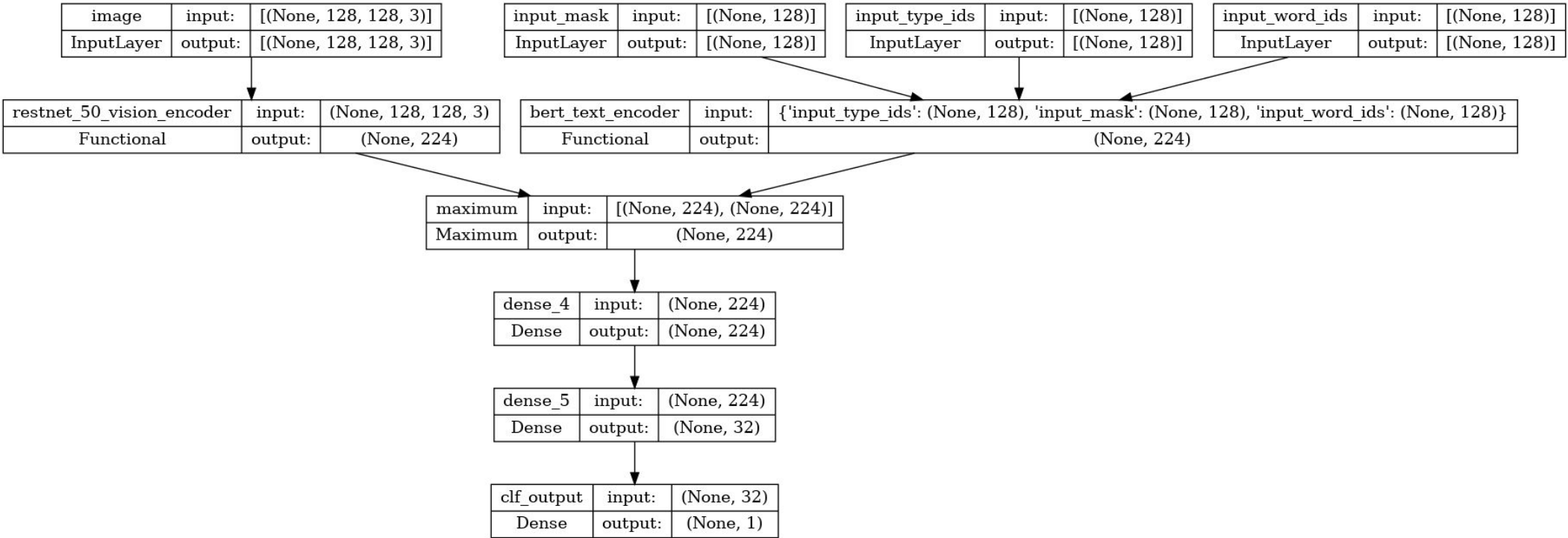


Current Model and Model Performance

Out-of-Sample Binary Classification Performance: True vs Fake

AUC	Sensitivity (0.5 cutoff)	Specificity (0.5 cutoff)
0.90	0.80	0.85

Model architecture: (30 + million parameters)



Next Steps

1. Model compression (primarily via quantization)
2. Container orchestration via Vertex AI pipeline
3. Build model endpoint
4. Build frontend UI

Questions?