

Ankit Badatala

703-989-0542 | ankitbadatala2019@gmail.com | ankitsdesk.vercel.app | github.com/asbadatala

EXPERIENCE

- Strategy** May 2025 – Present
Senior AI Engineer, Core AI team
- Co-engineered an enterprise AI product enabling businesses to upload over 100GB of data and use a chatbot to derive product/customer insights.
 - * Architected ReAct and ReWOO frameworks, developed 15+ specialized agents and MCP tool-calling.
 - * Optimized the RAG pipeline, reducing hallucinations 25% and improving Time to First Token 20%.
 - * Helped drive adoption of the AI product from 0 → 25 companies (e.g., Cisco, Pfizer) in 3 quarters.
 - Built Voice AI for Strategy mobile app serving 4,000 DAU.
 - * Built production-scale, streaming inference on AWS SageMaker to serve a custom Text-to-Speech model with < 500ms Time to First Audio.
 - * Improved GPU utilization and optimized KV cache to achieve 8× cost reduction (~87.5%).
 - Led a global team of 15 to build a computer vision pipeline that converts hand-drawn sketches into presentation ready dashboards.
 - * Trained RF-DETR object detection model achieving COCO mAP@0.50:0.95 = 0.55.
 - * Built a feature extractor + RBF-SVM binary classifier achieving ROC-AUC 1.0 and F1 0.984
 - * Automated color extraction from brand logos using MiniBatchKMeans to improve customer personalization.
- JPMorgan Chase** Aug 2023 – May 2025
Software Engineer, Personalization and Insights team
- Architected and deployed a solution to enforce a large scale data retention policy, impacting over 100,000 Chase customers, by using Terraform, Python, AWS, and Spinnaker
 - Built a python application to optimize AWS EMR cluster runtimes and pinpoint over-fitted instance types, reducing data processing costs by \$120,000 a year
 - Built a python application to identify and clean over 400 terabytes of noncurrent version data in AWS S3 buckets, saving \$72,000 a year in cloud storage costs
 - Orchestrated over 20 ETL pipelines in Control M for processing and storing petabytes of data

PROJECTS

AI-powered Portfolio Website | ankitsdesk.vercel.app

- Built a **RAG-powered chatbot with real-time Voice AI** to make my resume interactive. It can accurately answer any career related question, project specifics, and behavioral questions with my true experiences.
- Implemented smart chunking and semantic search with **Upstash Vector Database** and session memory with **Redis** to maintain conversation context and reduce hallucination.

TECHNICAL SKILLS

LLM: RAG, multi-agent frameworks (ReAct, ReWOO), prompt engineering, vector search & embeddings

Computer Vision: Object detection and image classification (YOLO, RF-DETR), hyperparameter optimization

Voice AI: Speech-to-text, open source Text-to-speech models, ultra low latency streaming

ML Frameworks & Tooling: PyTorch, TensorFlow, scikit-learn, ONNX, vLLM

Cloud & MLOps: AWS SageMaker, Terraform, GPU inference optimization

Certifications: AWS Cloud Practitioner

EDUCATION

Georgia Institute of Technology *In progress: Master of Science in Computer Science, Machine Learning*

Ohio State University *Bachelor of Science in Computer Science and Engineering, Machine Learning*