Vedant Tibrewal

213-696-0171 | veda.tibrewal@gmail.com | linkedin/vedant-tibrewal | github/vedant-tibrewal | vtibrewal.com | Los Angeles, CA

PROFESSIONAL SUMMARY

AI Engineer with 3+ years shipping production LLM pipelines, agentic systems, and CV models. Built an 8-stage multi-LLM fact-verification agent (KEPLER), a stateful LangGraph full-stack app (Archon), and a YOLOv5 edge model on a live warehouse forklift. ML at scale: 13K SKUs, 200GB+ on GCP. MS Applied Data Science, USC (GPA 3.9/4.0) · *F-1 OPT / STEM-OPT · Open to relocate*

SKILLS

LLM / Agentic: LangChain / LangGraph · MCP Servers · OpenAI (GPT-4o/4.1/5.1) · Anthropic APIs · RAG / pgvector · Structured Outputs · Agentic Pipelines · Prompt Engineering

ML & CV: PyTorch · TensorFlow · XGBoost / scikit-learn · YOLOv5/v6 · OpenCV · Transformers · Few-shot / TTT · mAP / RMSE / precision-recall

MLOps & Cloud: Vertex AI Pipelines · GCP (Composer, Dataproc, Cloud Build) · AWS (ECS Fargate, S3, CloudWatch) · Airflow · Docker · CI/CD (GitHub Actions, Jenkins) · Celery + Redis

Data & Backend: PySpark · Databricks · Snowflake · PostgreSQL / BigQuery · FastAPI / Django · SQLAlchemy · ETL / KPI Design

Languages & Cert: Python (primary) · SQL (advanced) · JavaScript / TypeScript · Java · **Google Cloud Certified — ACE** (2023)

EXPERIENCE

Tiger Analytics — *Machine Learning Engineer* Remote · Aug 2021 – Jul 2024

— Perfect Pallet: Computer Vision for Warehouse Automation

Tech: *Python, YOLOv5/v6, OpenCV, PyTorch, Edge Deployment (Intel CPU, 5-Camera)*

- Trained and deployed **YOLOv5 pallet integrity model (mAP@0.5 ≈ 0.76)** on edge device on an active forklift — no GPU; owned full pipeline (dataset curation, annotation QA, evaluation, real-time inference); replaced manual process addressing millions in annual warehouse losses.

— Price Elasticity Model: SKU-Level Pricing Intelligence

Tech: *Python, GCP Composer/Airflow, Dataproc, Vertex AI Pipelines, BigQuery, Cloud Build, Docker*

- **Productionized large-scale ML pipeline on GCP** (Airflow + Dataproc + Vertex AI); 4-week cadence across **∼13K SKUs, 13 categories** for Fortune 500 retail client; CI/CD via Cloud Build + Jenkins; automated executive reports with elasticity bucket classifications per run.

— Profit Pool: Global Profitability Analytics Platform

Tech: *Python, PySpark, Databricks, GCP Composer/Airflow, Snowflake, Power BI*

- **Automated ∼3-day manual cycle to ∼2 hours** via Airflow DAGs and data-quality gates; processed **∼26.7M records (>200GB)** with PySpark + Databricks; governed KPI layer (NSV, MAC, Margin) with auditable bridge-mapping, exception workflows, and run-level lineage.

Enverus — *Business Technology Analyst, Intern* Remote · Feb – Jul 2021

Tech: *Python, OCR, Object Detection, Layout Analysis, SQL*

- Built **PDF extraction pipeline for oil & gas receipts** — OCR + object detection converting scanned documents into structured database records; QA checks (missing-field rate, format validation) reduced manual overhead.

RESEARCH

Research Assistant, IMSC Lab — USC *(Computer Vision)* Aug 2024 – Present

Tech: *Python, PyTorch, YOLOv5/v6, OpenCV, IR+RGB Sensor Fusion*

- Investigating **IR + RGB object detection ensembling** for robustness under occlusion, lighting variance, and adverse weather; built custom evaluation harness with YOLO-style reports (mAP, precision/recall, per-class).

PROJECTS

KEPLER — Multi-LLM Agentic Fact-Verification Pipeline *(Collaborative)* **2024–2025**

Tech: *Python, OpenAI (GPT-4o/4.1/5.1), Anthropic APIs, Google Search API, Pydantic, Async Scraping*

- **8-stage agentic pipeline**: Claim Decomp → Evidence Retrieval → Reranking → Multi-LLM Verification → Confidence Scoring; ∼11s E2E; multimodal inputs; consensus across GPT-4o/4.1/5.1 + Claude Opus/Sonnet/Haiku with context-aware domain-authority weighting.

Archon — Stateful LangGraph AI App & LLM AutoFiller — Chrome Extension **2025**

Tech: *FastAPI, LangGraph, Celery, Redis, PostgreSQL, Next.js, SSE, JWT · React, Chrome MV3, IndexedDB, OpenAI/Anthropic APIs*

- **Archon:** Full-stack AI app (FastAPI + Next.js) with **stateful LangGraph workflow** — structured outputs, retry loops, quality-validation checkpoints; async Celery + Redis; live SSE streaming.
- **AutoFiller:** Chrome Extension auto-filling **50+ form types in <3s** via DOM extraction across LinkedIn, Greenhouse, Lever, Workday — zero backend; all data local (IndexedDB + chrome.storage).

EDUCATION

University of Southern California — MS, Applied Data Science Aug 2024 – May 2026

GPA: **3.9 / 4.0** · Coursework: Applied NLP, ML for Data Science, Data Mining, Information Retrieval

VIT Vellore — B.Tech, Electronics & Communication Engineering Jul 2017 – Jun 2021

GPA: **8.85 / 10** · Publication: *Identifying Stuttering Using Deep Learning*, IJITEE Vol. 8, 2019 · Patent **2020041022886** (Granted)