

# SIDDHARTH SHAH

shahsid2002@gmail.com · +1 814-321-6934 · [linkedin.com/in/siddharth29shah](https://www.linkedin.com/in/siddharth29shah) · [www.siddharthgshah.com](http://www.siddharthgshah.com) · New York City, NY (Open to Relocation - Dubai, Europe, Singapore)

AI/ML engineer & data scientist completing a U.S. B.S. at Penn State (Aug 2026) — 130+ credits across Computer Engineering, Applied Data Science & AI/ML. Delivered production-grade ML pipelines, cloud ETL systems, and LLM-powered applications across 3 professional roles. Seeking a Summer 2026 AI/ML, Software Engineering, or Data Analytics internship with particular interest in financial data infrastructure, quantitative analytics, and fintech ecosystem at the Dubai International Financial Centre, Dubai.

## EDUCATION

### Pennsylvania State University

Aug 2021 – Aug 2026 · State College, PA, USA

Bachelor of Science, Information Sciences & Technology (Minors: Computer Engineering, Applied Data Sciences — 130+ credits completed in-residence at University Park)

**Relevant Coursework:** Machine Learning · Database Systems · Distributed Networks · Statistical Modeling · Human-Centered Design & Development · Cybersecurity · Software Engineering · Linear Algebra

## PROFESSIONAL EXPERIENCE

### University Research Associate | ReSense Lab — Penn State University

Oct 2024 – Aug 2025 · University Park, PA

- Engineered real-time geospatial reasoning algorithms in Python, cutting driver-hazard false-positive rate by **15%** in simulation — directly improving model precision for V2V safety-critical scenarios.
- Mined **15,000+** vehicle telemetry records with Pandas & NumPy clustering to classify 8 high-risk driving event categories, establishing quantitative safety thresholds adopted into the lab's evaluation framework.
- Rebuilt low-latency V2V communication and Matplotlib visualization pipelines, reducing end-to-end system response time by **20%** and enabling real-time QA feedback loops for the research team.

### Data Analyst Intern | Matriosh

Jun 2024 – Aug 2024 · New York City, NY

- Automated multi-source ETL pipelines in Azure Databricks and Splunk across 5+ enterprise log streams, eliminating **40%** of manual data-processing effort and reducing mean time-to-incident-discovery.
- Designed 4 executive-facing Power BI and Tableau dashboards monitoring service KPIs, compliance SLAs, and network traffic anomalies — accelerating root-cause analysis cycles by an estimated **30%**.
- Generated strategic market intelligence via Python web scraping and SQL across **3 data sources**, producing insights that directly shaped quarterly business strategy and streamlined reporting workflows.

### Business Analyst & Applications Developer Intern | AllianceTek Inc.

Jun 2023 – Aug 2023 · Malvern, PA

- Re-architected RESTful Flask APIs and optimized React/TypeScript frontend components, delivering a **30%** reduction in page-load time and API latency — measurably improving end-user experience.
- Led requirements-gathering across 3 sprint cycles with cross-functional stakeholders, translating business processes into Axure RP wireframes and reducing design rework by **25%**.
- Applied Pandas and PostgreSQL analytics to cluster feature-usage patterns across **2 product lines**, surfacing data-backed insights that directly influenced roadmap prioritization and pricing strategy.

## PROJECTS

### Complag — AI-Native Fintech Compliance Platform

Dec 2025 – Present

[NestJS](#) · [FastAPI](#) · [LLM/RAG](#) · [pgvector](#) · [Knowledge Graphs](#) · [Chrome Extension](#)

- Architected a knowledge-graph + LLM-RAG backend for automated regulatory filing and real-time trade surveillance; a companion Chrome Extension cut form-submission time by **40%** and surfaced 25% more high-risk compliance events during QA.

### Predictive Cardio Health Monitoring System

May – Aug 2023

[TensorFlow](#) · [Scikit-learn](#) · [Flask](#) · [REST API](#) · [Automated Data Pipelines](#)

- Deployed a real-time cardio-risk prediction API with automated data-cleaning pipelines, achieving **65%** classification accuracy and reducing analyst workload by 30% versus the manual review baseline.

### AI Forensic Image Analysis Tool

Sep – Dec 2022

[CNN](#) · [TensorFlow](#) · [Keras](#) · [Deep Learning](#) · [Large-Scale Image Datasets](#)

- Led 10-engineer team to deliver a deep-learning tamper-detection and object-recognition pipeline; achieved **77%** detection accuracy and cut investigation triage time by 35% on large-scale datasets.

## TECHNICAL SKILLS

**AI / ML:** Supervised learning, classification & clustering, deep learning (TensorFlow, Keras), NLP (NLTK), LLM/RAG agent workflows, model evaluation & experimentation, big data (Spark / Hadoop familiarity), Quantitative analysis, Financial data pipelines, Compliance monitoring, KPI dashboard design

**Data & Analytics:** Pandas, NumPy, Scikit-learn, Power BI, Tableau, Excel; ETL pipeline design, Azure Databricks, Azure Data Factory, Splunk, PostgreSQL, MySQL, NoSQL

**Engineering:** Python, TypeScript / Node.js, Java, C++, Go, SQL; Flask, FastAPI, NestJS, React; Docker, CI/CD, Git; RESTful API design, systems integration, QA

**Languages:** English (Fluent) · Hindi (Native) · Gujarati (Native) · French (Conversational) · Sanskrit