

# YASH SHAH

[Portfolio](#) | [Linkedin](#) | [Github](#) | [yshah43@asu.edu](mailto:yshah43@asu.edu) | 480-952-2965

## EDUCATION

**Master of Science in Information Technology (GPA: 4.0/4.0)**

Arizona State University

**May 2025**

Arizona, USA

**B.Tech in Computer Engineering (GPA: 9.0/10.0)**

Charotar University of Science and Technology

**April 2023**

Gujarat, India

## EXPERIENCE

**IT Business Analyst Intern**

**June 2025 - Present**

Zydus Pharmaceuticals USA Inc.

New Jersey, USA

- Develop a multi-agent AI assistant in Microsoft Teams that routes sales queries through modular LLM and SQL validator agents to fetch Microsoft Fabric data, reducing manual report lookup time across sales operations.
- Enforcing security by Azure AD auth and agent-level SELECT-only filtering to keep queries safe and SLA-reliable.
- Streamline reporting with USPS-normalized PySpark addresses, fuzzy-matching strategies to achieve over 93% linkage across 600K+ records, align 25K+ mismatches, and validate results using OpenAI for quality assurance.
- Reconcile 1M+ SAP transactions using conditional logic to fix customer data for compliance and reporting.
- Built a Python pipeline to process 100+ drug PDFs with Camelot and pdfplumber for analysis in Microsoft Fabric.

**Data Engineer**

**January 2024 - May 2025**

Arizona State University

Arizona, USA

- Cleaned and standardized Excel-based student records for downstream Salesforce ingestion, then built a WorkOS UI using GraphQL + REST APIs to pull registrant counts and generate reports, reducing turnaround time by 52%.
- Contributed 10+ new feature implementations for the Smart DART app, leveraging React, TypeScript, Node.js, and applied AWS Lambda and DynamoDB for distributed processing, achieving a 2.6% latency reduction.
- Built Flask-Airtable inventory system to track 500+ SKUs, resolving sync delays and reducing manual workload.

**Software Engineer Intern**

**June 2024 - August 2024**

New York City Department of Finance

New York, USA

- Scripted modular SharePoint web parts utilizing React, SPFx, and Fluent UI with SPHttpClient integration to render list data and support metadata-based filtering, improving document retrieval turnaround by up to 35%.
- Deployed a React-based intake form with Formik and Yup for responsive form checks, routed data to SharePoint through Power-Automate, and enforced SharePoint permissions to maintain secure access control.
- Cleaned 50K+ records in Python and auto-mapped them to SharePoint via Power Automate with lossless transfer.

**AI Engineer**

**December 2022 - April 2023**

AI Genesis Glib

Gujarat, India

- Built Rasa conversational agent with YAML-based intents, entity extraction, and FastAPI endpoints to automate flight booking, seat selection, and payment tasks, reducing manual support escalations.
- Migrated NLP stack to LangChain and Pinecone, improving routing to ~95% intent and ~97% response accuracy.
- Validated system through pytest integration tests and log analysis, refining routing logic across 100+ scenarios.

## PROJECTS

**Coffee AI Agent ([Github](#)) | Python, LLaMA 3.1, Pinecone, React Native, Firebase, RunPod**

**August 2025 - Present**

- Design a modular AI chatbot using LLaMA 3.1 and agent based architecture (guard, classifier, recommender, order handler), deploying on RunPod GPUs to enable real time interactions with 3x lower response latency.
- Build RAG pipeline with OpenAI embeddings, Pinecone, and React Native-Firebase app for grounded LLM queries.

**Book Recommendation System ([Github](#)) | Python, LLMs, LangChain, Chroma, Gradio**

**May 2025 - July 2025**

- Implemented zero-shot classification with pretrained LLM embeddings to semantically encode 5,200+ book descriptions, achieving 78% fiction/non-fiction accuracy through cosine similarity based inference.
- Trained transformers on 7 emotion sentiments, achieving 66% accuracy and generating top-16 recommendations.

**YouTube Trending Video Analysis ([Github](#)) | Python, AWS, SQL, ETL**

**January 2024 - April 2024**

- Implemented an ETL pipeline on AWS (S3, Lambda, Glue, Athena) to ingest, transform, and analyze 100,000+ YouTube videos from multiple global regions, converting raw JSON into Parquet format for scalable analysis.
- Established an Amazon QuickSight dashboard to visualize key video metrics and highlight engagement trends.

## TECHNICAL SKILLS

**Languages & Frameworks:** Python, SQL, R, JavaScript, TypeScript, Express.js, Node.js, React, PowerShell

**AI/ML & NLP:** Transformers, Hugging Face, LangChain, Scikit-learn, OpenAI API, LLaMA

**LLMs & Vector DBs:** FAISS, Pinecone, Chroma, Rasa

**Model Ops & Deployment:** FastAPI, Flask, Streamlit, Gradio, Docker, Redis, Celery, REST APIs

**Cloud & BI Tools:** AWS, GCP, Microsoft Fabric, Azure, Snowflake, Airflow, Power BI, Tableau, QuickSight, Looker Studio