

Pranav Panchal

📍 USA ✉️ pranavpanchal1999@gmail.com ☎️ +1 (858) 603 1826 [in LinkedIn](#) 🌐 [Portfolio Website](#)

SUMMARY

Full Stack AI & Data Engineer with proven success building and scaling AI-driven products from the ground up at fast-moving startups. Expert in transforming messy, real-world data into structured insights and automated workflows, shipping robust features across the stack—from **scalable data pipelines and LLM-powered agents to React/TypeScript frontends and real-time APIs**. Hands-on with **Python, TypeScript, AWS, Databricks, Docker, and CI/CD**, I thrive in high-ownership environments, moving quickly from prototype to production. I've led projects like AI data agents, real-time voice automation, and no-code analytics, reducing manual effort by up to 90% and driving 3x faster analytics. Ready to join a small, ambitious team and build impactful systems.

SKILLS

Certifications: [Azure Databricks Platform Architect](#), [IBM AI Engineering](#), [AWS Cloud Practitioner](#), [Python](#), [MongoDB](#)
AI & LLMs: OpenAI API, Prompt Engineering, LangChain, LangGraph, RAG, Code Generation, LLM Debugging, ChromaDB
Data Engineering: ADF, Databricks, Microsoft Fabric, Delta Lake, Spark, Kafka, Snowflake, Fivetran, Unity Catalog Data
Data Analysis: Python (Pandas, NumPy, PySpark), SQL, pandasql, Data Profiling, Anomaly Detection, Data Validation
Big Data Technologies: Kafka, Spark, Hive, Apache Airflow, Redshift, MongoDB, Docker, Kubernetes, dbt
Machine Learning & Forecasting: Regression, Classification (Decision Trees, Random Forest, Naive Bayes), Forecasting, Predicting
Visualization: Power BI, Tableau, Plotly Dash, Matplotlib
Web/API: FastAPI, REST, JWT Auth, SQLAlchemy, Celery
Cloud & DevOps: AWS (ECS, RDS, S3, Lambda, DynamoDB), Azure, Vercel, Docker, Kubernetes, Terraform

PROJECT

DataSage – AI-Powered Interactive Data Notebook

May 2025

- Architected asynchronous **agent orchestration using LangChain, FastAPI, ChromaDB**, enabling real-time automated analysis and accelerating workflow execution by 80%.
- Enabled natural language query-to-code and dashboard automation via **OpenAI LLMs**, democratizing self-service analytics.
- Automated **data validation, profiling, and anomaly detection in Python**, elevating dataset quality and reducing manual review.
- Deployed a modern **React/TypeScript** interface with one-click **Plotly visualization**, driving high user adoption.
- **Tech Stack:** React, TypeScript, FastAPI, Python, Pandas, Plotly, Matplotlib, OpenAI API, LangChain, ChromaDB, RAG

Healthcare AI Clinical Assistant

June 2025

- Developed a real-time clinical voice agent using **OpenAI GPT, Whisper, and RAG**, automating live transcription, **STT/ASR**, and SOAP note creation.
- Integrated **vector search** for instant retrieval of patient history, boosting provider efficiency and decision accuracy.
- Automated **PHI redaction and error logging in Python**, improving documentation integrity and compliance.
- **Tech Stack:** Python, FastAPI, Hugging Face Transformers, OpenAI GPT/Whisper, RAG, VectorDB

AI Audiobook Platform

July 2025

- Designed a scalable backend in **Python and FastAPI** for rapid manuscript-to-audio conversion, supporting multi-character narration with **ElevenLabs and OpenAI TTS**.
- Developed **RESTful APIs** and asynchronous processing with **Celery**, ensuring efficient audio generation and secure data handling.
- Designed secure, multi-user **React frontend**; managed **JWT auth** and async job processing; optimized **file storage on AWS**.
- Enabled instant audio preview and feedback via a **React frontend**, enhancing author and publisher experience.
- **Tech Stack:** Python, FastAPI, Celery, PostgreSQL, ElevenLabs API, OpenAI TTS, Azure AI Foundry, Docker, JWT Auth, SQLAlchemy, React

EXPERIENCE

AI Data Engineer

Sansar Tec

Aug 2024 – Present, Remote, USA

- **Built self-healing data pipelines capable of processing and validating millions of records**, with automated error tracking and **Slack alerts** to minimize downtime and accelerate incident response.
- Delivered a 60% reduction in manual effort and 3x faster analytics by deploying **AI-powered workflow** assistants and automation.
- Implemented **LLM-powered agents and Retrieval-Augmented Generation (RAG)** to automate data cleaning, analysis, and dynamic dashboard creation, reducing manual data prep time by 70% and accelerating insight delivery for business stakeholders.
- **Automated live dashboards using Dash Plotly and deployed them to production**, enabling real-time insights for stakeholders.
- Increased deployment speed by 20% and pipeline reliability by enhancing **CI/CD automation with Docker and Kubernetes**.
- Reduced data loading time for 1,000 JSON files from 4 hours to 22 minutes by **optimizing pipelines in Azure Data Factory and Databricks with parallel processing and schema validation**.
- Automated JSON flattening and Shopify data ingestion in **Python**, cutting manual effort by 90% and ensuring consistent data quality.

Pricing Analyst

Sally Beauty Supply LLC <ul style="list-style-type: none">• Built SQL-based data pipelines and automated ETL with Alteryx for daily pricing and vendor data integration.• Developed and maintained pricing models in Python to optimize pricing strategies across 5,000+ SKUs.• Identified \$30M+ in vendor pricing discrepancies through SQL analysis, supporting contract renegotiations.• Built Tableau dashboards to visualize SKU trends and pricing anomalies, improving audit readiness.	Jun 2023 – Aug 2023, Denton, USA
Data Engineer Global Vox <ul style="list-style-type: none">• Built scalable data pipelines in Databricks and AWS S3, automating 5M+ records weekly and cutting manual workload by 70%.• Automated ETL workflows using PySpark, SQL, and AWS Lambda, improving data delivery speed by 60% for real-time analytics.• Integrated AI/ML models (regression, classification, anomaly detection) to enhance data integrity and reduce compliance issues by 50%.• Developed backend APIs and monitoring dashboards with Power BI, Azure Monitor, and AWS CloudWatch, cutting incident response time by 50%.• Built forecasting models in Python, boosting resource utilization by 30% and ensuring 95%+ on-time delivery.• Standardized data models and automation, decreasing support tickets by 60% and enabling self-serve analytics for 20+ analysts.• Documented backend, AWS, and ML workflows, reducing onboarding time by 40% and ensuring audit compliance.	Jul 2020 – Jun 2022, Ahmedabad, India
Python Developer Intern Brainy Beam Technologies <ul style="list-style-type: none">• Developed Python scripts to automate repetitive ETL tasks, improving workflow efficiency and saving 10+ hours/week• Developed data ingestion scripts to extract and process raw datasets from CSV and Excel files, enabling faster data onboarding for business use cases.• Built custom Python utilities to validate, clean, and standardize large datasets, improving downstream data quality and reliability.	May 2020 – Jul 2020, Ahmedabad, India

EDUCATION

Master of Science, Business Analytics (Data Engineering Track) University of Texas at Dallas	Aug 2022 – May 2024, Dallas, USA
Bachelor of Technology, Information Technology Indus University	Jul 2017 – May 2021, Ahmedabad, India