

# SHREYA JAIN

+1 (734)450-9984

✉ [shreyadjain26@gmail.com](mailto:shreyadjain26@gmail.com)

🌐 [LinkedIn](#)

📁 [Portfolio](#)

🐙 [GitHub](#)

## EDUCATION

### University of Michigan - Ann Arbor

Master of Science in Data Science

May 2025

Ann Arbor, MI

### Savitribai Phule Pune University

Bachelor of Technology in Information Technology

May 2023

Pune, MH, India

## WORK EXPERIENCE

### AI Research Engineer

Mar 2025 – Present

Michigan Medicine

Ann Arbor, MI

- Deployed (Streamlit/AWS) multi-domain RAG Engine with **FAISS retrieval** across ASO, CRISPR, & Global Health
- Integrated Semantic Scholar **API retrieval**, improving model accuracy up to 21% overall and **46% on the positive class** on rare-variant classification.
- **Benchmarked 300 gene variants with 2400+ predictions**, on two exon-skipping strategies (ES-RF & ES-VE) with a **resumable, rate-limit-aware** harness and provenance **auditing**.
- Cut token costs by 60% through query-aware model routing and improved ES-VE eligibility accuracy from 0.60 to 0.81.
- **Evaluated 300 variants** against **4 models** and ground truth using **precision, recall, F1 score, and confusion matrices**.

### Research Assistant

May 2024 – Dec 2024

University of Michigan

Ann Arbor, MI

- Built a Python/NLP extraction pipeline combining rule-based parsing, schema-driven processing, and custom **spaCy NER** (7 domain labels) for historic texts.
- Fine-tuned **GPT-based NER and Llama-3-8B** with **PyTorch, Hugging Face, LoRA**, and quantization, reaching 98% extraction accuracy, 75% domain entity recognition, and **66% F1** on specialized entities.
- Wrote comprehensive **unit tests** and automated validation for outlier detection, extracted outputs, and consistency checks, **reducing validation time by two hours per dataset**.

### AI & Software Intern

Oct 2022 – Jul 2023

DRDO

Pune, MH, India

- Led UI dev. for a product, from **Figma** prototyping through the development and deployment of a **Flask** application.
- Reduced pose-keypoint jitter by 35% while validating 5 Army drill postures using **MediaPipe, OpenCV**, and Python.
- Enabled real-time feedback by classifying 15+ keypoints using joint-angle calculations, visibility, and live error feedback.
- **Achieved 98% detection accuracy** and reduced background noise by engineering a **YOLOv8/PyTorch** auto-cropping pipeline, deployed as an offline **Flask** application for government use.

### Data Science Intern

Aug 2022 – Dec 2022

Beyond Business Travel

Ireland

- Saved 8+ staff-hours daily by replacing manual invoice entry with a **React** and **Django** document-automation website.
- Achieved **98% extraction accuracy** using **Python, pypdf, OCR**, layout-aware parsing, and automated validation.

## PROJECTS

### FoodLink [GitHub](#) – Generative AI, Supply Chain, REST APIs, Next.js, TypeScript, Prisma, MapLibre, Claude API.

- Led a hackathon team and built a food-bank inventory and logistics platform for a supply chain hackathon using **Next.js, React, TypeScript, Prisma, and SQLite**.
- Integrated Anthropic Claude multimodal AI to extract structured inventory data from images, emails, CSVs, and text.
- Developed 14-day inventory **forecasting** for shortages, expiring stock, and inbound/outbound shipments.
- Created a **MapLibre geospatial** marketplace for surplus redistribution, transfer tracking, and driver coordination.

### RoleSave - Job Application Tracker [GitHub](#) – TypeScript, Next.js, React, Supabase, PostgreSQL, Docker

- Built a full-stack platform that tracks jobs and preserves JD through a **Chrome extension** and **Next.js** dashboard.
- **Automated email classification** and confidence-based matching for confirmations, interviews, offers, and rejections.
- Engineered a secure, idempotent processing pipeline with **PostgreSQL RLS**, background queues, **Docker-isolated** PDF conversion, retries, and **41 automated tests**.

## TECHNICAL SKILLS

**Languages:** Python, R, SQL, C, C++, JavaScript, TypeScript, HTML, CSS **Frameworks:** React.js, Node.js

**AI/ML:** PyTorch, TensorFlow/Keras, Scikit-learn, Hugging Face, LangChain, RAG, LLMs, Generative AI, Causal Inference

**Databases:** Relational Databases: SQL, MySQL **NoSQL:** MongoDB

**Statistics:** Statistical Inference, Hypothesis Testing, ANOVA, Time-Series (ARIMA), Predictive Modeling

**Data/Viz:** Pandas, NumPy, SciPy, NetworkX, Tableau, Power BI, Matplotlib, Seaborn, Altair, GeoPandas

**Developer Tools:** Git, GitHub, Copilot, Figma, Docker, Streamlit, GCP