

Shaunak Rahul Mahajan

+1 934-263-0993 | shaunakmahajan.datasci@gmail.com | [linkedin.com/in/shaunak-mahajan](https://www.linkedin.com/in/shaunak-mahajan) | [GitHub](https://github.com/shaunakmahajan) | shaunakmahajan.pages.dev

EXPERIENCE

Splashgain Technologies

Pune, India

AI Software Engineer

Aug 2023 – Mar 2024

- Built and deployed a production **Conversational System** for **Eklavya** and **e-Pravesh**, automating support for **500+ global customers** and **5,000+ concurrent users** on enterprise platforms processing **30M+ assessments** across **15+ countries**, reducing repetitive support requests.
- Redesigned **AngularJS** components and data-loading workflows for **Eklavya**, improving UI responsiveness by **35%** under peak examination traffic through efficient data loading and responsive rendering, keeping high-concurrency assessment sessions stable for simultaneous users.
- Implemented **SQL Stored Procedures** for backend retrieval and reporting workflows, aligning retrieval logic with frequently accessed examination datasets to improve database access efficiency and response times for assessment reporting and administrative operations.
- Developed **Highcharts** monitoring dashboards for grades, comparative rankings, and student-performance analytics across **30M+ assessments**, consolidating examination results into interactive real-time views that accelerated access to operational insights.

Splashgain Technologies

Pune, India

Software Intern

Jul 2022 – Jul 2023

- Engineered a real-time **OpenCV computer-vision pipeline** processing webcam streams at **20+ FPS** to detect unauthorized **phones and earphones** during online examinations, automating proctoring without interrupting live sessions across varied webcam hardware.
- Fine-tuned an **EfficientDet-D0** object-detection model with **TensorFlow** on custom annotated datasets, validating predictions across varied lighting conditions and webcam hardware to reliably identify unauthorized objects during real-time examination proctoring.
- Analyzed recurring support-query patterns and built an automated **chatbot workflow** for common user requests, translating repeated request patterns into automated flows that reduced manual responses and laid the foundation for the production conversational system.

Mahindra and Mahindra

Nashik, India

Data Analytics Intern

Aug 2021 – Sep 2021

- Trained an unsupervised **LSTM autoencoder** on **41,639 CNC servo-speed readings** and evaluated **9,432 held-out test readings**, flagging **48 anomalies** whose MAE reconstruction error exceeded a **0.7056 threshold** set from training loss.
- Ingested a **4GB CNC telemetry dataset** into **MongoDB** through chunk-based **Pandas** processing without loading the full dataset into memory, enabling scalable querying and analysis of **112K+ servo-speed observations** for downstream signal preprocessing.
- Built a time-series preprocessing pipeline with **Pandas**, **NumPy**, **MongoDB**, and **linear interpolation**, filling missing telemetry values and converting irregular signals into cycle-aligned sequences using CNC main-program execution intervals as consistent model inputs.

PROJECTS

AI-Powered Application Outreach Assistant [Github]

Sep – Oct 2026

- Built a **React/TypeScript** and **FastAPI** outreach assistant integrating **Tsenta MCP**, **Serper**, and **Ollama** to research contacts and draft grounded LinkedIn notes across **797 tracked applications**, with durable **LangGraph** checkpoints and validation gates for restart recovery.
- Implemented **SQLAlchemy** and **PostgreSQL/pgvector** retrieval with embedding caches and lexical fallback, enforcing **200-character** notes and session/CSRF protections, with draft editing and provider-failure recovery covered by **130 pytest** and **11 Playwright** tests.
- Deployed the app as a non-root **Docker** container on **AWS EC2** via **CloudFormation** with **Secrets Manager**, encrypted storage, and zero inbound ports, gated by **GitHub Actions CI** running backend, browser, and container smoke tests on every push.

RAG Hallucination Detector [Github] [Model] [Dashboard]

Oct 2026

- Fine-tuned **ModernBERT-base** in a custom **PyTorch** loop on **100K RAGTruth** sentence pairs to flag unsupported claims in **LLM-generated RAG** answers, reaching **0.65 F1** and **0.92 AUROC** on held-out test vs 0.20 F1 and 0.64 AUROC for a zero-shot NLI baseline.
- Trained with **fp16 autocast**, gradient accumulation, token-budget batching, and class-weighted loss on a **Kaggle T4**, then reached **71.5 balanced accuracy** on **11 LLM-AggrFact** datasets and cut calibration error from **0.043 to 0.016** via **Platt scaling**.
- Served per-sentence calibrated scores through a **FastAPI** endpoint in a non-root CPU **Docker** image, gated by **GitHub Actions CI** running **46 pytest** tests and a container smoke test, and published results to a **Tableau Public** dashboard.

Efficient Test-Time Scaling via Execution-Supervised Refinement [Github]

Sep 2025

- Achieved **88.3% MBPP**, **98.3% GSM8K**, and **76.7% BoolQ** accuracy across **150 sampled problems** with Gemini 2.0 Flash at **1.35 model calls** per MBPP problem, beating the MiPS best-of-32 baseline of 57.8% on MBPP with over **95% fewer calls**.
- Architected **Self-Refine-Loop**, a **Gemini 2.0 Flash** pipeline that generates code, runs unit tests in an isolated **Python** subprocess with a 10-second timeout, and feeds error traces such as **NameError** back for up to **3 refinement iterations**.
- Doubled MBPP accuracy from **40.0% to 88.3%** by **Context Injection** of missing helper classes such as **Pair** and **Node** into the prompt and execution scope, after an ablation traced most failures to undefined data structures rather than wrong logic.

TECHNICAL SKILLS

LLM & Agents: LLMs, RAG, AI Agents, LangGraph, Prompt Engineering, LLM Evaluation, MCP, Embeddings, Gemini API, Ollama

Languages & Data: Python, TypeScript, JavaScript, SQL, PostgreSQL, pgvector, MongoDB, SQLAlchemy, Pandas, NumPy

Web & APIs: FastAPI, Pydantic, React, Vite, Django, AngularJS, REST APIs, Highcharts, Tableau

Machine Learning: PyTorch, Hugging Face Transformers, NLP, TensorFlow, Scikit-learn, LightGBM, LSTM, OpenCV

Cloud & Tools: Docker, AWS (EC2, CloudFormation, Secrets Manager), GitHub Actions, Git, pytest, Playwright, Claude Code

EDUCATION

Stony Brook University

New York, USA

Master of Science in Data Science

Aug 2024 – May 2026

Vishwakarma Institute of Technology

Pune, India

B.Tech in Information Technology

Aug 2019 – May 2023

PUBLICATIONS

Effective Mid-Day Meal Analysis Depending on Children's Needs using Machine Learning, Grenze Int. J. Eng. Technol., 2023 [Paper]

Detection and Categorization of Patients Having Alzheimer's Disease: Machine Learning Algorithms, AIP Conference Proceedings, 2023 [DOI]

Brain Tumor Classification using Machine Learning, Journal of Pharmaceutical Research International, 2021 [DOI]