

Swaraj Shaw

Machine Learning Engineer • NLP • Speech • LLM Systems
Dublin, Ireland • +353 89 984 9430 • shaw.swaraj16@gmail.com
swarajshaw.github.io • linkedin.com/in/swarajs • github.com/swarajshaw
Work Authorisation: Stamp 4 (Ireland – No sponsorship required)

Summary

Machine Learning Engineer with 5+ years of experience delivering NLP, Speech (TTS/ASR), and LLM-based systems across Big Tech and startups. Expert in text-to-speech modelling, embeddings, ONNX inference, Python/Rust ML infrastructure, evaluation design, and large-scale data quality automation. Strong track record building high-impact AI platforms, offline inference engines, ML pipelines, and production-grade products.

Technical Skills

- **ML/AI:** NLP, LLMs, ASR/TTS, embeddings, G2P, TN rules, evaluation, hyper parameter tuning
- **Cloud/Infra:** AWS (SageMaker, Lambda, Textract), GCP (Vertex, BigQuery), Docker, Terraform
- **Frameworks:** PyTorch, TensorFlow, ONNX Runtime, HuggingFace, Scikit-learn
- **Data Engineering:** SQL Server, PostgreSQL, MySQL, ETL, Data Lakes
- **Backend/Systems:** Rust, Python, C++, FastAPI, Next.js, Celery, Redis
- **Tools:** spaCy, SearXNG, Tauri, Git, Linux, Power BI, Tableau, Looker

Experience

Meta (via Covalen) Dublin, Ireland
Text-to-Speech Linguist (Hindi Locale) Oct 2023 – Present

- SME for Hindi TTS: phonesets, G2P mappings, TN rules, and full linguistic pipeline design.
- Built a **14k+ sentence evaluation suite** across 11 linguistic categories; automated coverage analysis.
- Developed large-scale lexicon QA tools; corrected **14k+ phonetic errors** in a 35k-word database using Python.
- Achieved **95.6%** benchmark accuracy through iterative training, error analysis, and pipeline refinement.
- Authored extensive annotation, TN, and G2P guidelines—standardised contributions across linguists.
- Major contributor to Ray-Ban Meta smart glasses Hindi TTS: evaluated **30k+** scripts, normalised data, and short-listed **200+** voices.
- Collaborated with research/engineering teams on training platform integration (Bento, MLHub) and quality tooling.

AspireNXT Pvt. Ltd. Singapore (Remote/Hybrid)
Business Intelligence & ML Engineer Jan 2020 – Jan 2022

- Built conversational chatbots (Amazon Lex + Connect) across Slack, Messenger, and internal tools.
- Designed an NLU pipeline (Lex + Dialogflow) improving routing and reducing manual workload.
- Delivered a **recommendation engine** using Amazon Personalize + SageMaker, increasing conversions by **15%**.
- Built anomaly detection pipelines for energy data; reduced operational wastage by **15%**.
- Automated BI dashboards (Amazon QuickSight, Tableau); improved data visibility and onboarding KPIs by **50%**.

Zeo Minds IT Solutions Hyderabad, India
Research Data Scientist May 2019 – Jan 2020

- Built credit-risk and churn models using SMOTE + XGBoost; improved accuracy by **15%**.
- Prototyped ASR-based booking and customer support workflows integrated with ML APIs.

Digital Nest Pvt. Ltd. Hyderabad, India
Data Analyst Jun 2018 – Apr 2019

- Built collaborative filtering recommendation models improving sales by **23%**.
- Created CNN-based image classifiers achieving **95%** accuracy for internal automation use cases.

E Info Solutions Pvt. Ltd. Kolkata, India
Android Developer Sept 2017 – Jun 2018

- Led Android app for Quest Mall; achieved **5,000+** installs and increased footfall via location-intelligent features.

AI Platforms & Engineering Work

Nexus AI Engine — End-to-End AI Platform (DataPrep Monorepo)

Private Repo

- Architected a **full-stack AI platform monorepo** combining FastAPI ingress, Next.js messenger UI, async workers, GPU inference, dataset builders, distillation flows, and evaluation pipelines under one modular workspace.
- Designed a production-grade **FastAPI gateway** for routing, prompt orchestration, rate limiting, tenant separation, search augmentation, dataset logging, and redaction-aware request processing.
- Built a **Next.js 19 messenger UI** (LLM chat, audio support, Launchpad experiment portal, scheduler UI) with model cards, history, and experiment controls.
- Implemented fully automated **dataset export pipelines** (LLM JSONL + Whisper CSV): deterministic PII redaction, audio caching, redaction logs, and dataset snapshot registration for lineage.
- Added support for **teacher–student distillation**: queue difficult samples, call local/remote teacher models, generate refined training pairs, and integrate distilled output back into dataset snapshots.
- Integrated **SearXNG**-backed web search with reranking, markdown extraction, and time-range filters, enabling high-quality verified answers in the AI Lab playground.
- Built a modular **MLOps layer with Prefect** for nightly dataset exports, distillation cycles, evaluation runs, and automated registry publishing.
- Created a **Launchpad metadata service** (FastAPI) storing experiments, runs, manifests, evaluation scores, and model lineage; added “Promote to Playground” gating logic after evaluation passes.
- Implemented **AI Lab training pipelines**: QLoRA fine-tuning for LLMs, Whisper adaptation, evaluation harnesses (BLEU, EM, WER), and model registry publishing (S3-compatible artifact store).
- Added **infrastructure templates**: Docker Compose (API + Worker + Redis), Terraform-ready GPU provisioning hooks, and Vast.ai automated GPU bringup scripts.

Sorted — Local-First AI File Intelligence System (Rust, ONNX, Tauri)

Private Repo

- Built an offline semantic file intelligence app with Rust daemon + GPU-accelerated ONNX C++ inference (Metal/DirectML).
- Implemented recursive scanners, duplicate detection, ignore filters, and SQLite metadata store (WAL mode).
- Designed embedding pipelines (MiniLM/BGE/Gemma) enabling semantic renaming, folder recommendation, and deduplication.
- Developed a Tauri desktop UI: previews, batch actions, and undo-safe operations backed by SQLite audit trails.
- Designed multi-stage ML experimentation (Python) + inference integration (C++/Rust) inside one monorepo.

Job Tracker Portal — ML-Powered Hiring Intelligence Platform

Private Repo

- Built end-to-end automated hiring cockpit: job scraping, normalization, ML extraction, feedback loops.
- Implemented MiniLM-based FastAPI ML extraction service; dashboard auto-fills JD fields.
- Added full retraining workflow: edited corrections → JSONL dataset → embeddings → updated model.
- Engineered modular backend (Node.js + PostgreSQL) with impersonation, gated access, and auditing.

Document Extractor (OCR + ML Pipeline)

- Automated document intelligence system using Textract + PyTorch; reduced manual workload **25%**.

Visual Product Recommender

- Large-scale image similarity recommender (SageMaker + Elasticsearch) increasing revenue by **30%**.

Education

Dublin Business School — MSc in Artificial Intelligence (1.1 Honors)

2023

Amity University — B.Tech in Computer Science and Engineering

2019

Languages

English (Fluent), Hindi (Native)