

RISHET MEHRA

rishetmehra11@gmail.com | +91 7982913080 | LinkedIn | GitHub

Education

B.Tech, Mathematics and Computing

2023 – 2027

Delhi Technological University (DTU)

Experience

AI Research & Development Intern

Jul 2025 – Sep 2025

Zanista AI (UK)

- Built Python scrapers with **Scrapy**, **Playwright**, and **Selenium** to pull **JS-rendered pages** behind paywalls and **login/auth** flows, handling rate limits, retries, and **HTTP status-code** edge cases to ingest **10,000+ articles** daily.
- Wrote **regex** and custom **parsers** to extract clean fields from messy **HTML** and inconsistent **JSON**, then **deduplicated** and **normalized** the data into **PostgreSQL**; schema validation and retry logic cut manual cleanup to near zero.
- Built **LLM pipelines** (summarization, NER, sentiment, topic clustering) on the cleaned corpus, cutting manual review time and sharpening downstream data quality.

UI/UX & Web Development Intern

Nov 2025 – Jan 2026

BharatFare Travel & Technologies (UK)

- Triaged a **281-issue backlog** and shipped **40+ fixes** across every core user flow, cutting page load times and stabilizing critical paths under a large, messy queue.
- Ran a competitor analysis that surfaced 3 UX gaps, all adopted into the product roadmap.

Projects

LeadPilot

Python, Scrapy, BeautifulSoup, Apify, Pandas

- Built a **multi-source scraping pipeline** pulling business data from **Google Maps**, company websites, and **Instagram** by category and location, using **regex** to extract emails and contact fields from raw **HTML**.
- Normalized 5,000+ heterogeneous records** into a single structured dataset with **Pandas**, **deduplicated** and enriched them, then scored leads for outreach.

Codex-spend

Node.js, SQLite, SQL, npm, CLI

- Published an **npm CLI** that parses OpenAI Codex's local **SQLite** databases with **SQL queries** to surface token-burn patterns and accurate cost estimates factoring in prompt-caching discounts.
- Parsed session data from both **SQLite** and streamed **JSONL** logs, handling multiple Codex **storage-format variations** to **normalize** token usage into a single consistent cost report.

VidWise AI

Python, DuckDB, FAISS, LangChain, RAGAS, Streamlit

- Built a **retrieval and processing pipeline** over YouTube transcripts (**FAISS** retrieval, context compression, MultiQuery), logging evaluation metrics to **DuckDB** and tracking retrieval drift on a live **Streamlit** dashboard.

SpaceATC

Python, LangGraph, FastAPI, PyTorch, YOLO

- Built a **multi-agent system** coordinating satellite collision-avoidance maneuvers with a central lowest-fuel selector gated by human-in-the-loop approval, plus an onboard CV module estimating 6-DOF pose feeding a deterministic safety check.

Skills

Languages: Python, C++, JavaScript, TypeScript, SQL

Data Engineering: Web scraping (Scrapy, Playwright, Selenium, BeautifulSoup), regex, HTML/JSON/PDF parsing (PyPDF2), OCR (Tesseract), data cleaning, normalization, deduplication, PostgreSQL, SQLite, REST APIs, HTTP/auth, Pandas

AI/ML: LangChain, LangGraph, RAG, FAISS, LLM APIs, PyTorch, HuggingFace, multi-agent systems, LLM evaluation

Full-Stack & Tools: React, Next.js, FastAPI, Flask, Node.js, Docker, Git, Cursor, Claude Code

Achievements

- Rank 27 / 16,000+, Google PromptWars Virtual (national Google AI competition).
- 5x National Hackathon Winner.
- Published 2 open-source npm packages on the npm registry.
- Co-Head, E-Cell DTU (500+ members); organized E-Summit '25 (40,000+ footfall).