

Harsh Kashyap

Meerut, India | 📞 +91 7351534994 | ✉️ harshkashyap664@gmail.com
🌐 [linkedin.com/in/harsh-kashyap-bioinformatics](https://www.linkedin.com/in/harsh-kashyap-bioinformatics) | 🐙 github.com/Imharsh08

Professional Summary

Data Engineer, AI Practitioner, and Full-Stack Developer with 4+ years of experience building production data pipelines, LLM-powered applications, relational database systems, and enterprise workflow automation. Hands-on expertise across the full Generative AI stack — from prompt engineering and agentic workflow design to AI-assisted automation and analytics integration. Proficient in Python, SQL (PostgreSQL/MySQL), React, and Bash, with strong software engineering fundamentals including modular design, version control, and documentation. Holds an M.Sc. in Bioinformatics & Computational Biology, adding domain fluency in life sciences data. Adept at translating complex technical capabilities into measurable operational outcomes for both technical and non-technical stakeholders.

Technical Skills

Languages	Python, JavaScript (React, Next.js, TypeScript), SQL, R, HTML/CSS, Bash
Databases & SQL	PostgreSQL, MySQL, Supabase, Airtable; schema design, query optimisation, indexing, ETL
Data Engineering	Pipeline design & automation, event-driven systems, workflow orchestration, scalable data processing, data consistency
Generative AI & LLMs	Anthropic Claude API, ChatGPT, Google Gemini; prompt engineering, agentic workflow design, LLM pipeline integration; actively developing with LangChain/LangGraph
AI Productivity Tools	GitHub Copilot, Cursor, Replit AI, n8n, Make (Integromat), Zapier, Notion AI, Monday AI; AI-assisted coding and automation
NLP & ML	BERT (semantic similarity, sentiment analysis), NLP pipelines, recommendation systems, fraud detection; Python (scikit-learn, transformers)
No-Code / Low-Code	AppSheet, Google Apps Script, Power Automate, Airtable Automations, Zapier
BI & Reporting	Power BI (DAX), Tableau, Looker Studio, Google Sheets, Excel
Software Engineering	Git/GitHub, Unix/Linux, Docker (working knowledge), modular architecture, technical documentation
Cloud & Platforms	Supabase, Google Cloud, Cloudflare Pages, Microsoft Azure (awareness)
Domain Knowledge	Life Sciences data (Genomics, Biopharma context); biological databases (NCBI, UniProt, RNACentral); M.Sc. Bioinformatics

Professional Experience

MIS / Data Analyst & Full-Stack Developer — Mangla Rubber Technologies Mar 2025 – Present
Partapur, India

- Built and deployed a **full-stack React + PostgreSQL (Supabase) platform** for enquiry tracking, quotation generation, and order automation — a live production system serving real-time operational data.
- Designed relational database schemas and **optimised SQL queries** for performance under concurrent production load; implemented secure access via Google Cloud Authentication.
- Architected **agentic, event-driven automation pipelines** using AppSheet and Google Sheets, eliminating manual workflows and reducing operational turnaround time by **60%**.
- Replaced legacy tracking systems with structured **relational workflows**, improving data integrity and accessibility across cross-functional teams.
- Built **AI-assisted reporting dashboards** in Power BI and Looker Studio with advanced DAX measures, cutting reporting latency and enabling real-time leadership decision-making.
- Led **digital transformation and stakeholder training** initiatives, improving team operational efficiency by **80%** across supply chain operations.

Freelance Data Analyst — Excise Department, Uttar Pradesh
Uttar Pradesh, India

Apr 2020 – Mar 2024

- Managed and optimised **SQL database systems** consolidating multi-sector operational data, eliminating 85% of physical paperwork.
- Designed and deployed **8+ dynamic dashboards** across 6 sectors, improving district-wide data visibility and reporting accuracy by 70%.
- Applied **statistical modelling for demand forecasting** and inventory optimisation, reducing carrying costs by 88%.
- Collaborated with government stakeholders to **define KPIs, gather requirements, and deliver data-driven insights** for operational decision-making.

Data Science Intern — Oasis Infobyte

Oct 2023 – Nov 2023

Remote

- Built end-to-end **Python ML pipelines** for sentiment analysis (85% accuracy), recommendation systems, and fraud detection.
- Applied NLP techniques — text preprocessing, feature extraction, model evaluation — across multiple production-grade prototypes.

Bioinformatics Intern — BioTecNika Info Labs Pvt. Ltd.

May 2023 – Jul 2023

Bengaluru, Karnataka, India

- Used **generative AI and ML tools for scientific literature curation**, cutting manual research time by 50%.
- Contributed to structural bioinformatics workflows, improving biomolecular structure prediction success rate by 15%.

Key Projects

ProSched — Production Planning Web Application

Jun 2025 – Present

Next.js, TypeScript, PostgreSQL, GitHub

- Built a full-stack production scheduling app with **Gantt charts, weekly planning views, and relational filtering** for manufacturing workflow visualisation.
- Clean modular architecture, Git version control, documented API design — production-ready engineering standards.

Dispatch Email Automation Pipeline

Dec 2025 – Present

AppSheet, Google Sheets, Google Apps Script, HTML

- Designed an **event-driven pipeline** triggering automated, branded HTML emails on dispatch confirmation; reduced communication latency from hours to seconds with zero manual intervention.
- End-to-end pipeline ownership: trigger logic, data transformation, templating, and delivery.

AI-Powered Virtual Assistant for Scientific Database Navigation

M.Sc. Thesis

Python, Botpress, LLM Integration, NCBI / UniProt

- Built an **LLM-integrated agentic assistant** enabling researchers to query biological databases (NCBI, UniProt) via natural language — an early example of agentic AI workflow design.
- Integrated ML-based query understanding and intent classification; aligns with human-in-the-loop AI curation use cases.

Semantic Similarity Prediction Model

Academic Project

Python, BERT, NLP

- Implemented a **BERT-based semantic similarity model**; optimised text preprocessing and embedding workflows for high-accuracy contextual understanding.

Employee KPI Automation Dashboard

Professional

Google Apps Script, Google Sheets

- Scripted end-to-end automated KPI tracking pipeline — data ingestion, computation, scheduling, and dashboard delivery — demonstrating full pipeline ownership.

Education

M.Sc. in Bioinformatics & Computational Biology

2021 – 2023

Ch. Charan Singh University, Meerut, India | 77% — Rank Holder

B.Sc. in Mathematics

2018 – 2021

Meerut College, Ch. Charan Singh University, Meerut, India

Certifications & Training

- **Bacterial Genome Assembly and Quality Control** — ECDC (European Centre for Disease Prevention and Control), Dec 2023
- **CRISPR Workshop & Advanced Microscopy Techniques in Life Science** — BioTecNika Info Labs, 2023
- **Data Science** — Oasis Infobyte, Sep 2023
- **Introduction to SQL for Data Science** — Scaler, Jul 2023
- **Tableau Specialization** — Jobaaj.com, Oct 2023
- **Advanced MS Excel** — Jobaaj.com, Jan 2024