



# DHAIRYA PAREEK

Data Engineer · AI Builder

Berlin, Germany (Open to Relocation)

 [dhairya-pareek.site](https://dhairya-pareek.site)  [github.com/happyhnyb](https://github.com/happyhnyb)

 [linkedin.com/in/dhairya-pareek](https://linkedin.com/in/dhairya-pareek)  [dhairya@hnyb.in](mailto:dhairya@hnyb.in)

## PROFESSIONAL SUMMARY

Data Science & AI student pursuing **bachelors from GISMA University of Applied Sciences, Berlin** with 10+ months of industry experience building production data systems. Experience in end-to-end data engineering — from ingestion and pipeline design to analytics dashboards and AI-powered workflows. Proven track record of working across backend API integrations, Spark-based big data processing, Snowflake/Databricks warehouse architectures, and applied machine learning. **Seeking data engineering or AI systems internships across Europe, India, and Remote.**

## KEY HIGHLIGHTS

- 6+ production-grade projects shipped across data engineering, applied ML, and analytics domains.
- 10 months of real industry experience delivering backend integrations and data infrastructure in a US product company.
- Hands-on with enterprise-grade tools: Snowflake, Databricks, Apache Spark — covering Indian and global tech stacks.
- Available for immediate internship start — Europe, India, and Remote opportunities.

## EXPERIENCE

### Product Development Engineer

Mar 2024 – Jan 2025 · 10 months

**Choice 1 Communications (C1C)** · Tampa, FL · Remote

- Shipped production backend API integrations with GoHighLevel (GHL), including OAuth auth flows, webhook orchestration, and bi-directional data synchronisation across CRM and marketing systems.
- Designed and deployed end-to-end data pipelines for ingestion, transformation, and reporting — reducing manual data processing time by ~60% across marketing and operational workflows.
- Built production-grade analytics dashboards aggregating multi-source campaign performance data, enabling real-time business metric visibility for stakeholders.
- Developed automation system integrations that streamlined lead workflow management, cutting manual intervention across key operational processes.
- Contributed to AI-powered analytics experimentation — prototyping workflow intelligence features using structured prompt engineering and output validation.

## PROJECTS

### Formula One Analytics Platform · Python · Apache Spark · Databricks · Snowflake · Looker Studio

- Flagship end-to-end analytics platform ingesting 75 years of F1 race data. Built Python ETL pipelines, Spark transformation layers on Databricks, Snowflake warehouse models, and Looker Studio BI dashboards — covering full data lifecycle from raw ingestion to business insight delivery.

### **GA4 Analytics Dashboard Pipeline** · [Google Analytics 4](#) · [Python](#) · [SQL](#) · [Looker Studio](#)

- End-to-end analytics pipeline that extracts GA4 event data via API, applies Python-based transformation logic, models it in SQL, and surfaces decision-ready marketing performance dashboards — closely mirroring production analytics stacks at Indian product startups.

### **AI Studio — Reinforcement Learning Environment** · [PyTorch](#) · [TensorFlow](#) · [Python](#)

- Custom RL training environment with reward shaping, multi-episode training loops, and experiment tracking. Implemented policy gradient methods and Q-learning variants with structured evaluation workflows.

### **High-Frequency Trading Simulation Engine** · [Python](#) · [NumPy](#) · [Quantitative Modelling](#)

- Python-based simulation framework for backtesting trading strategies under synthetic market conditions, modelling order book dynamics, latency effects, and strategy performance at tick-level resolution.

### **Large-Scale Movie Data Platform on Databricks** · [Databricks](#) · [Apache Spark](#) · [PySpark](#) · [SQL](#)

- Distributed data processing pipeline on Databricks for large-scale ingestion, cleaning, and aggregation of movie metadata. Leveraged PySpark for parallelised transformations across multi-GB datasets.

### **Automated Data Mining & Ingestion System** · [Python](#) · [Web Scraping](#) · [Data Pipelines](#) · [Automation](#)

- Scalable scraping and ingestion system with validation, deduplication, and structured output for downstream analytics. Designed for reliability with configurable retry logic and schema enforcement.

## **TECHNICAL SKILLS**

---

**Data Engineering:** Apache Spark, PySpark, Databricks, Snowflake, ETL/ELT Pipelines, dbt (familiar), Airflow (familiar)

**Programming:** Python, SQL, JavaScript, R

**Cloud & Platforms:** AWS (S3, Glue), GCP (BigQuery, GA4), Databricks, Snowflake

**AI / ML:** PyTorch, TensorFlow, Reinforcement Learning, Supervised/Unsupervised ML, Prompt Engineering

**Analytics & BI:** Looker Studio, Power BI, Google Analytics 4, A/B Experimentation

**Data Stores:** PostgreSQL, MongoDB, Snowflake

**Tools:** Git, REST APIs, Webhooks, Docker (familiar), VS Code, Jupyter

## **EDUCATION**

---

**B.Sc. Data Science & Artificial Intelligence**, GISMA University of Applied Sciences, **Berlin, Germany** · **2023 – Present (4 Year Course)** · Focus: Data Engineering, Machine Learning, Applied AI Systems

---

[dhairya-pareek.site](#) · [linkedin.com/in/dhairya-pareek](#) · [github.com/happyhnyb](#)