Khushi Malik

Delhi – 110007 | khushimalik511263@gmail.com | GitHub | LinkedIn | HuggingFace | Portfolio

EDUCATION

- **Kirori Mal College, University of Delhi** — Aug 2023 – May 2027
  B.Sc. Mathematics (CGPA: 8.27) — *Delhi, India*

WORK EXPERIENCE

- **Planto AI** — May 2025 – Present
  Lead AI Researcher (Aug 2025 – Present) AI Researcher (May 2025 – Jul 2025) — *Remote, New Delhi*
  - Led RAG pipeline optimization, tokenization & prompt compression research, HuggingFace model fine-tuning, technical presentations, and AI-powered VS Code extension development team
  - **Tech:** Python, HuggingFace, LangChain, RAG, Vector Databases, LLM Fine-tuning

- **Zindi** — Apr 2025 – Present
  Freelance Data Scientist — *Remote, South Africa*
  - Participated in competitive AI challenges achieving Top 10 in Amini GeoFM crop classification, handling data preprocessing, model development, and leaderboard optimization across agriculture and healthcare domains.
  - Skills: Data Science, Geospatial AI, Clinical Reasoning, ML Model Optimization

- **Unified Mentor** — Apr 2025 – May 2025
  Data Analyst Intern — *Remote, Delhi*
  - Extracted job market trends from 2,000+ listings, built Random Forest salary prediction model, and delivered insights through visual storytelling.
  - Skills: Random Forest, Data Visualization, Statistical Analysis

PROJECTS

- **Codex-Finetune — Intelligent Coding Copilot** — Jun 2025
  Repo/Model
  - Fine-tuned `starcoder2-3b` on CodeSearchNet with **LoRA**; instruction-tuned for multi-task gen (code, docstrings, funcs).
  - Efficient HF pipelines with Accelerate; tracked **3 epochs/3k steps** on W&B; final loss ∼1.20.
  - Instruction-tuned model; API-ready via Transformers.

- **AI Doctor Assistant — Fine-Tuned LLaMA 8B** — May 2025
  GitHub
  - Specialized clinical model using DeepSeek-R1-Distill-LLaMA-8B + **QLoRA** (Unsloth, 4-bit).
  - Trained on **500 samples / 60 steps**; **CoT prompting** improved explainability.

- **AI Coder Pro — VS Code AI Extension** — Jul 2025
  GitHub
  - Role: Developer; Tech: TypeScript, Node.js, VS Code API, Together AI.
  - Multi-agent tools: code review, bug detection, refactoring, test generation.
  - Chat UI with memory + file upload; privacy-first; compared favorably with Copilot/TabNine.

- **Tourist AI (In Progress)** — Aug 2025 – Present
  Personal Project
  - Building an **AI-powered tourist app** using **RAG** and open-source API models.
  - Goal: Smart itinerary Q&A from local knowledge; offline-first design.

ACHIEVEMENTS AND CERTIFICATIONS

- **Top 10 (World)** — Amini GeoFM AI/ML Hackathon (crop-type classification with satellite data). [Certificate]
- Excel: Business Intelligence — Codebasics [Credential]
- SQL: Beginner to Advanced — Codebasics [Credential]
- Python: Beginner to Advanced — Codebasics [Credential]
- EV Sales Analysis — Codebasics [Certificate]

SKILLS

- **Languages:** Python, SQL, Java, JavaScript, TypeScript
- **AI/ML:** Scikit-learn, TensorFlow, PyTorch, Hugging Face, OpenAI API, Transformers, LangChain, OpenCV, PEFT
- **Domains:** LLMs, Generative AI, NLP, Computer Vision, RAG, Fine-tuning, Prompt Engineering
- **Tools/Cloud:** Git, Docker, VS Code, AWS, GCP, Excel, Power BI, MySQL, Jupyter, MongoDB, Firebase
- **Web/APIs:** Flask, FastAPI, Node.js, Express.js, React, HTML/CSS, Postman
- **Analytics:** Pandas, NumPy, Data Cleaning, ETL, Statistical Analysis, Data Visualization
- **Mathematics:** Probability, Statistics, Linear Algebra, Calculus

EXTRACURRICULAR

- **DAKSHI** (College Research Society): Co-Head, Report Compilation & Data Management Committee; Head of LinkedIn Page.
- **NCC Cadet** — served through Aug 2025.
- Open-source contributions and participation in AI competitions; active GitHub portfolio.