

# KHUSHI MALIK

DELHI – 110007

• [khushimalik511263@gmail.com](mailto:khushimalik511263@gmail.com)

• [linkedin.com/in/Khushi-malik-analyst](https://www.linkedin.com/in/Khushi-malik-analyst)

• [Portfolio](#)

## Experience

### Lead AI Researcher

(Aug 2025 – Present)

### AI Researcher

(May 2025 – Jul 2025)

Planto AI · Remote | New Delhi, India

- Led a team in developing **AI Coder** – a VS Code extension integrating AI-powered code assistance features.
- Conducted research on **Retrieval-Augmented Generation (RAG)**, **tokenization strategies**, and **text-to-prompt compression** to optimize LLM performance.
- Delivered technical presentations on **RAG pipelines**, **prompt engineering**, and **token efficiency techniques**.
- Applied **practical fine-tuning** methods on Hugging Face models for domain-specific tasks using real-world datasets.

### Freelance Data Scientist

Apr 2025 – Present

Zindi · Remote | South Africa

- Solving real-world data science problems through global AI competitions.
- Performing end-to-end pipelines: data cleaning, feature engineering, and model deployment.
- Applied LLM-based techniques for boosting model accuracy using advanced augmentation methods.

### Data Analyst Intern

Apr 2025 – May 2025

Unified Mentor · Remote | Delhi, India

- Analyzed large-scale datasets using SQL and statistical modeling techniques.
- Built insightful dashboards and reports to support business decisions.
- Gained proficiency in 6+ tools including Excel, Power BI, Pandas, and NumPy.

## Projects

### Codex-Finetune — Intelligent Coding Copilot([GitHub](#))

Fine-tuned StarCoder2-3B for multi-task code understanding and generation

Jun 2025

**Tools:** Python · Hugging Face · PEFT (LoRA) · PyTorch · W&B

- Fine-tuned **bigcode/starcoder2-3b** on CodeSearchNet (Python) using LoRA for efficient training.
- Built an instruction-tuned model capable of code generation, docstring creation, and function completion.
- Designed memory-efficient data pipelines using Hugging Face **datasets** and **accelerate**.
- Tracked training (3 epochs, 3k steps) via Weights & Biases; achieved final loss ~1.20.
- Packaged model for inference with Hugging Face Transformers; integrated API-ready deployment.

### AI Doctor Assistant – Fine-Tuned LLaMA 8B Project ([GitHub](#))

May 2025

Built a specialized medical language model using **DeepSeek-R1-Distill-LLaMA-8B**, fine-tuned with LoRA (**Unsloth, 4-bit QLoRA**) for clinical reasoning tasks.

**Tools:** Transformers, PEFT, Unsloth, Hugging Face Datasets

- **Goal:** Simulated human-like diagnosis using Chain-of-Thought (CoT) prompting.
- **Training:** 60-step fine-tuning on 500 clinical samples (**medical-o1-reasoning-SFT**).
- **Result:** Improved accuracy and explainability in medical question-answering.

### AI Coder Pro – VS Code AI Assistant([GitHub](#))

Jul 2025

**Role:** Developer | **Tech:** TypeScript, Node.js, VS Code API, Together AI, Mocha, ESLint

- Built an open-source VS Code extension integrating **Together AI** for smart code assistance.
- Developed multi-agent tools for **code review**, **bug detection**, **refactoring**, and **test generation**.
- Designed a **chat-based UI** with file upload, persistent memory, and context-aware responses.
- Enabled **custom LLM integration**, focusing on transparency, privacy, and developer control.
- Compared favorably with GitHub Copilot and TabNine; fully extensible and free to use.

## Technical Skills

- **Tools:** Excel, Power BI, Power Query, MySQL, Jupyter Notebook
- **Languages:** Python, SQL
- **Libraries/Frameworks:** GitHub, Pandas, NumPy, Hugging Face, PEFT, Torch, TensorFlow, PyTorch
- **Visualization & Reporting:** Canva, PowerPoint, MS Word
- **Analytics & Modeling:** Data Analysis, Statistical Analysis, ETL, Data Cleaning, Regression Analysis.
- **Mathematics:** Statistics, Probability, Linear Algebra, Calculus

## Education

B.Sc.(Mathematics) | KiroriMalCollege, University of Delhi

Aug 2023 – May 2027

• CGPA: 8.27

north delhi-110007

## Extracurricular / Certifications

- Excel: Business Intelligence – Codebasics
- SQL: Beginner to Advanced – Codebasics
- Electric Vehicle Sales Analysis – Codebasics
- Python: Beginner to Advanced – Codebasics

[–View Credentials](#)

[– View Credentials](#)

[–View Certificate of Appreciation](#)

[- View Credentials](#)