

ANOOSHKA BAJAJ

✉ anooshkabajaj@gmail.com
in [linkedin.com/in/anooshkabajaj](https://www.linkedin.com/in/anooshkabajaj)

☎ (812) 822-7876
github.com/ab490

📍 Bloomington, IN (open to relocation)
🌐 ab490.github.io

EDUCATION

Indiana University Bloomington

Master of Science (M.S.) in Data Science

May 2026

GPA: 3.94 / 4.00

- **Courses:** Applied Machine Learning, Computer Vision, Big Data Management.

Indian Institute of Technology Mandi (IIT)

Bachelor of Technology (B.Tech.) in Bio-Engineering with Minor in Computer Science

May 2023

CGPA: 8.50 / 10.00

- **Courses:** Probability and Statistics, Data Structures and Algorithms, Deep Learning.

TECHNICAL SKILLS

Programming & Scripting: Python, C++, SQL, Bash

ML & AI Frameworks: PyTorch, scikit-learn, LangChain, Google ADK

Data & Visualization: PostgreSQL, MySQL, Power BI, Matplotlib, Plotly

Tools & Platforms: Linux, Git, Docker, Cloud (GCP, AWS), HPC

EXPERIENCE

Data Science Intern - TGS

Houston, TX

May 2025 - Aug 2025

- Optimized a **foundation model** pretrained on 1.1M+ geological samples through architectural modifications and latent-space analysis using t-SNE clustering to evaluate representation quality, and explored multimodal extensions using VLMs.
- **Fine-tuned** the model on 271k samples, improving performance on the downstream task of formation top prediction.

Research Assistant - Indiana University Bloomington

Bloomington, IN

Sep 2024 - present

- Researched in-context learning and interpretability in **Transformers**, using **NLP** techniques to analyze temporal effects in next-token prediction and uncovering biases including primacy, recency and contiguity, drawing parallels to human memory.
- Performed ablation studies on **LLMs** up to 9B parameters (Llama, Gemma) using **PyTorch** with multi-GPU parallelism.

Software Engineer - Radisys India Ltd.

Bangalore, Karnataka, India

Jul 2023 - Jun 2024

- Optimized **5G** OAM codebase in C++ and Python, reducing service latency across large-scale telecom network elements.
- Delivered feature upgrades for **CM** configuration module using 3GPP-compliant YANG configurations to increase throughput, and enhanced the **PM** performance module by designing metric counters, collaborating across teams in an **Agile** environment.

Data Engineer Intern - Technocolabs Softwares Inc.

Indore, Madhya Pradesh, India (Remote)

Jul 2022 - Sep 2022

- Preprocessed 1M+ records with Python and SQL for feature engineering and exploratory data analysis on large-scale datasets.
- Built a **BigQuery** data warehouse and automated ETL workflows by generating DAGs in **GCP (Google Cloud Composer)**, ensuring reproducible pipelines and delivering 20+ dashboards in **Google Data Studio** for real-time insights.

PROJECTS

Multi-Agent Race Telemetry Analysis | Google ADK, Vertex AI, FastAPI

- Built an **AI agent platform** for autonomous racing telemetry analysis using **Google ADK** and **Vertex AI** on GCP, developing agents that answer natural language queries over high frequency sensor data with a **FastAPI** backend and React dashboard.
- Designed a **multi-agent system** with three specialized agents for telemetry data discovery, natural language race analysis, and automated visualization, containerized with **Docker** and deployed on **Google Cloud Run**.

Enterprise RAG Copilot | RAG, LangChain

- Built a **Retrieval Augmented Generation** system using Gemini on **Vertex AI** and **LangChain** for citation-backed question answering using **FAISS** semantic vector search, supporting enterprise workflows and agent-based task automation.
- Developed a **FastAPI** backend for document ingestion, chunking, embedding pipelines, semantic retrieval and evaluation.

PUBLICATIONS

- Deven Mahesh Mistry, **Anooshka Bajaj**, Yash Aggarwal, Sahaj Singh Maini, Zoran Tiganj. “**Emergence of Episodic Memory in Transformers: Characterizing Changes in Temporal Structure of Attention Scores During Training.**” *Proceedings of the Nations of the Americas Chapter of the Association for Computational Linguistics (NAACL)*, 2025.
- **Anooshka Bajaj**, Deven Mahesh Mistry, Sahaj Singh Maini, Yash Aggarwal, Zoran Tiganj. “**Beyond Semantics: How Temporal Biases Shape Retrieval in Transformer and State-Space Models.**” *Proceedings of the 19th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, 2026.