Ankit Singh Chauhan

(317) 491-7832 | ankichau.1718@gmail.com | linkedin.com/in/chauhan-aankit

Skills

- **Languages**: Python, R, SQL, GO, C++, JavaScript, Java, TypeScript
- **AI/ML**: Scikit-learn, TensorFlow, PyTorch, HuggingFace, JAX, LangChain, OpenCV, Transformers, RAG, ChromaDB
- **Data Engineering**: Apache Spark (PySpark, MLlib), Kafka, HDFS, Hive, BigQuery, Real-time and Batch Pipelines
- **Cloud/MLOps**: Docker, Kubernetes, GitHub CI/CD, AWS, SageMaker , MLflow, Azure, GCP, CUDA, FastAPI, React
- **Databases**: MySQL, PostgreSQL, Cassandra, MongoDB, Redis, Elasticsearch, Neo4j
- **Experimentation**: Statistical Analysis, PowerBI, Experimental Design, A/B Testing, LLM Fine-Tuning, Model evaluation
- **Certifications**: Deep Learning Specialization, IBM Data Science Professional, Microsoft Certified: Azure AI Fundamentals

Work Experience

Indiana University Indianapolis **Aug 2025 - Present**
AI Engineer *Indianapolis, IN*

- Developed and containerized a full stack knowledge graph app with **36 endpoint API** (FastAPI), interactive graph visualizations (Cytoscape.js), and a responsive UI (React, TypeScript), enhancing the organization's ability to visualize complex community relationships effectively.
- Designed a Neo4j graph database schema (**10 node/11 relationship types**) based on an Ecological Systems Model, writing advanced Cypher queries for managing complex linking between people, programs, and assets, which improved data integrity by detecting and resolving orphans.
- Engineered a weak supervision pipeline and fine-tuned a BERT model to classify asset-deficit language, expanding a 270 example seed set by 5x into a **1,350 example corpus**.
- Deployed the fine-tuned BERT classifier (**0.68 F1 score**) as a containerized Flask app, optimizing the inference endpoint to a **p95 latency of 3.4s** under load.

Indiana University Indianapolis **Sep 2023 - May 2025**
Research Assistant *Indianapolis, IN*

- Led the research and development of **CATpc: Critical Activity Teacher Planning Companion** for an NSF funded research grant. The final system demonstrated **14% higher pedagogical alignment scores** and **8% reduced hallucinations** compared to the GPT-3.5 baseline.
- Built CoDe-KG, a knowledge graph extraction pipeline, by producing a sentence complexity dataset from 7,500 PubMed lung cancer abstracts, fine-tuning BERT variants on the dataset. Generated a labeled corpus by classifying on coreference resolved abstracts. Processed corpus with various prompting strategies (GIP, CoT, FICL, Hybrid) to perform sentence simplification and extract relationships (knowledge triples).
- Evaluated the triple extraction pipeline on standard relation extraction tasks, achieving **92.4% F1** on 398 gold standard triples. Outperformed previous methods with an **+8% macro-F1** gain on the ReBEL benchmark (**65.78%** vs previous ~57%).
- Supported over 30 students as a TA in the Fall 2024 H518 Deep Learning course through tutorials, labs, office hours, and lesson planning on topics like AlexNet, LSTM, Transformers, word2vec, BERT, GAN, and Reinforcement Learning.

Capgemini **Apr 2020 - Sep 2022**
Cloud Consultant (Data & Machine Learning Platforms) *Mumbai, Maharashtra*

- Engineered a data ingestion pipeline using PySpark to load data into over 100 Hive tables, **reducing data processing time by 40%**.
- Built an XGBoost based predictive forecasting model that optimizes SKU selection for enterprise licensing needs, delivering **18% annual cost savings** on underutilized license spend for 43,000+ users.
- Created efficient Cloud Functions to process and load over 10 GB of daily raw JSON data from GCS into BigQuery and Cloud SQL, enhancing data accessibility for business intelligence applications.
- Managed the successful migration of legacy applications to Azure cloud environment during an enterprise split, ensuring service continuity and operational efficiency for over 23,000 users.

Wipro Limited **Jul 2018 - Mar 2020**
Cloud Consultant (Cloud Infrastructure & DevOps) *Navi Mumbai, Maharashtra*

- Developed federated workflows for Azure infrastructure provisioning, **managing 250+ virtual machines** to ensure high availability and consistent uptime.
- Designed and maintained Azure DevOps pipelines for deployments to Azure App Services for 20+ business and operations applications, **cutting deployment time by ~30%**.
- **Authored 20 knowledge base documents** on routine administration tasks and contributed to regular knowledge sharing sessions on topics such as storage lifecycle, retention, and Azure CLI across EMEA, APAC, and NA teams.

Research and Academic Projects

Benchmarking LLMs for Pairwise Causal Discovery in Biomedical and Multi-Domain Contexts **Jan 2025 - May 2025**
Indiana University Indianapolis

- Managed the experimental code for evaluation for pairwise causal discovery (PCD), benchmarking **13 open-source LLMs** on **12 diverse datasets** using various prompting styles (few-shot, COT, FICL, ReAct, etc) and advanced span-scoring (Hungarian matching, SentenceLM similarity).
- Benchmarked detection and extraction capabilities of LLMs with DeepSeek-R1-Distill-Llama-70B attaining top **causal detection accuracy (49.57%)** and Qwen2.5-Coder-32B-Instruct led **causal-pair extraction (47.12%)**.
- Ensured robust evaluation (8 expert annotators, inter-annotator kappa of at least 0.758) and released a reproducible benchmark contributing to a publication on LLM's causal reasoning capabilities.

Cultural Eval: Quantifying Cultural Bias in LLMs (Independent Study) | Github **Jan 2025 - May 2025**
Indiana University Indianapolis

- Developed a quantitative framework to benchmark cultural bias in LLMs by applying **PCA to extract 5 latent cultural dimensions** from 97,000 records across 96 variables.
- Compared responses from Llama-2 13B, Gemma 3 12B, and Phi 4 across 5 extracted latent cultural dimensions using Tucker's Congruence Coefficient and Cohen's d, revealing consistent underestimation of Religious-Traditional values for non-Western demographic profiles (Cohen's d: -0.89 to -1.17).
- Created Western Bias Index and Overall Cultural Bias Index metrics, finding Llama-2 demonstrated the strongest Western bias **(WBI = 0.78)** and Phi 4 the highest overall cultural bias **(OCBI = 0.79)**.

Mod-Guide: An LLM-based Content Moderation Feedback System | Github **Aug 2024 - Jan 2025**
Indiana University Indianapolis

- Prototyped a multimodal NLP workflow parsing (OCR & Whisper) 200 community collected text, images, and audio documents of hate speech against minority voices in Bangladesh. Augmented the data with custom chain-of-thought reasoning, grounded in external knowledge sources, to generate contextual metadata.
- The final model's performance was evaluated on a 50 example held out set where our GPT-4 + RAG system achieved an **F1 score of 87%** at detecting culturally nuanced hate speech compared to an F1 score of 64% from the zero-shot GPT-4 baseline.

Symptom-Based Disease Prediction **Nov 2022 - Apr 2023**
University Of Mumbai

- Scraped medical symptom data and applied MeSH-based term normalization for semantic consistency.
- Developed an RNN classifier using Python and TensorFlow that **achieved 87% prediction accuracy** across 500 classes.

Education

Indiana University, Indianapolis **Aug 2023 - May 2025**
Master's, Applied Data Science

- **GPA:** 3.87/4
- **Coursework:** Statistical Learning, Deep Learning, Computer Vision, Database Design, Cloud Computing

University Of Mumbai **May 2020 - May 2023**
Bachelor, Information Technology

- **GPA:** 9.1/10
- **Coursework:** Operating Systems, Python, OOPM, Data Structures, DBMS, Java, Artificial Intelligence

Leadership and Awards

Indiana University Indianapolis **Aug 2024 - May 2025**
Treasurer (Graduate and Professional Student Government)

- Elected as the treasurer of Graduate Professional Student Government (GPSG) to represent student interests and **manage a budget of $85,000.**
- Served as co-chair to IU Funding Board, **oversaw allocations of $230,000 in grants** to organizations.
- Recipient of the **Graduate Student Honors Award** for Academic Excellence while maintaining leadership role.