Ankit Singh Chauhan

Indianapolis, IN 46202, United States | (317) 491-7832 | ankichau.1718@gmail.com | LinkedIn | Github

Education

Indiana University, Indianapolis **Aug 2023 - May 2025**
Master's, Applied Data Science
- **GPA:** 3.87/4
- **Coursework:** Statistical Learning, Deep Learning, Computer Vision, Cloud Computing, Informatics Research

University Of Mumbai **May 2020 - May 2023**
Bachelor, Information Technology
- **GPA:** 9.1/10
- **Coursework:** Digital Electronics, Embedded Systems, Python, Java, OOPM, DS, Operating Systems, Computer Graphics, IoT, AI

Publications

- Sydney Anuyah, Sneha Shajee Mohan, Bofu Dong, **Ankit-Singh Chauhan**, Sunandan Chakraborty, "Benchmarking LLMs for Pairwise Causal Discovery (PCD) in Biomedical and Multi-Domain Contexts,".2025 IEEE International Conference on Big Data (To appear).
- Dipto Das, Achhiya Sultana, **Ankit Singh Chauhan**, Saadia Binte Alam, Mohammad Shidujaman, Sunandan Chakraborty, and Syed Ishtiaque Ahmed. "Can LLM-based Content Moderation Identify Insensitive Speech toward Indigenous Ethnic and Religious Minorities?,".ACM Journal on Computing and Sustainable Societies (Under review).

Academic Projects and Research

CultureEval: Quantifying Cultural Alignment in LLMs (Independent Study) **Aug 2024 - May 2025**
Indiana University Indianapolis *Indianapolis, IN*
- Constructed a quantitative framework to measure cultural alignment and bias in LLMs using principal component analysis (PCA) over survey data from ~97k respondents across 96 sociocultural indicators, yielding 5 latent cultural dimensions.
- Evaluated Llama-2 13B, Gemma 3 12B, and Phi 4 along these latent dimensions using Tucker's Congruence Coefficient and Cohen's d, showing systematic underestimation of Religious-Traditional values for non-Western demographic profiles (Cohen's d: 0.89 to 1.17).
- Proposed Western Bias Index (WBI) and Overall Cultural Bias Index (OCBI) metrics; Llama-2 exhibited the strongest Western bias (WBI = 0.78), while Phi 4 showed the highest overall cultural bias (OCBI = 0.79).

Analyzing Chart-to-Text Dataset using Traditional CV (CNN-RNN) and VLM's **Aug 2024 - Dec 2024**
Indiana University Indianapolis *Indianapolis, IN*
- Conducted a comparative study of traditional CNN–RNN captioning models and modern VLMs on the Chart-to-Text dataset (27k+ Statista charts with associated metadata) to generate natural language chart summaries.
- Developed an encoder–decoder architecture replacing ResNet-50 encoder baseline with EfficientNet-B2 and a dual-LSTM decoder with coverage attention to avoid repetition, improving BLEU-4 from 0.18 to ~0.50.
- Investigated LLaMA 3.2 baseline with a LoRA fine-tuned LLaMA 3.2 Vision model and observed catastrophic forgetting. GPT as Judge score decreased from 0.56 to 0.36.

Research and Experience

Indiana University Indianapolis **Aug 2025 - Present**
AI Engineer *Indianapolis, IN*
- Developed and deployed a full stack knowledge graph app with 36 endpoint FastAPI backend, interactive Cytoscape.js graph visualizations, and a responsive React/TypeScript to support exploration of complex community ecosystems.
- Designed a Neo4j graph schema (10 node/11 relationship types) grounded in an Ecological Systems Model, and authored Cypher queries to model complex relationships among people, programs, and assets within community.
- Fine-tuned BERT to classify asset vs deficit language, expanding a 270 example seed set into a 1,350 example labeled corpus. Deployed the classifier as a containerized Flask app achieving F1 = 0.68, with a p95 inference 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 pedagogical chatbot. The system achieved a 14% improvement in pedagogical alignment and an 8% reduction in hallucinations relative to a GPT-3.5 baseline.
- Contributed to a knowledge graph extraction pipeline, by constructing a sentence complexity dataset from 7,500 PubMed lung cancer abstracts and fine-tuning BERT variants variants for complexity classification on coreference resolved abstracts. Applied multiple prompting strategies (GIP, chain-of-thought, FICL, hybrid prompts) to perform sentence simplification and relation extraction, generating knowledge triples for downstream graph construction.
- Benchmarked the triple extraction pipelines, achieving 92.4% F1 against 398 gold-standard triples and 65.78% macro-F1 on ReBEL, outperforming previously reported methods.

Indiana University Indianapolis **Aug 2024 - Dec 2024**
Teaching Assistant *Indianapolis, IN*

- Supported 30+ graduate students through weekly tutorials, labs, and office hours on convolutional networks (eg: AlexNet), recurrent models (LSTM), Transformers, word2vec, BERT, GANs, and introductory reinforcement learning.
- Assisted in designing assignments and in-class activities emphasizing model implementation details, empirical evaluation, and critical reading of deep learning research papers.

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

- Engineered data ingestion pipelines using PySpark to populate 100+ Hive tables.
- Built an XGBoost forecasting model to optimize SKU selection for enterprise software licensing, reducing under utilized license spend by 18% annually across 43,000+ users.
- Developed Cloud Functions to process and load over 10 GB of daily raw JSON data from GCS into BigQuery and Cloud SQL.
- Led migration for legacy applications to Azure during an enterprise split, ensuring service continuity for 23,000+ users.

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

- Designed and maintained CI/CD pipelines in Azure DevOps to automate testing and deployment of internal business and operations applications to Azure App Services.
- Authored 20+ knowledge-base documents on routine administration tasks and led regular knowledge sharing sessions on topics storage lifecycle management, retention policies, and Azure CLI best practices across EMEA, APAC, and NA teams.

Skills

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

Leadership and Awards

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

- Elected to represent graduate student interests and oversee financial operations for GPSG, managing an $85,000 annual budget.
- Served as co-chair of the IU Funding Board, overseeing allocation of $230,000 in grants to student organizations.
- Recipient of the Graduate Student Honors Award for Academic Excellence while holding a leadership role in student governance.