NAMAN AHUJA

(+1)602-369-9185 ◇ nahuja11@asu.edu ◇ linkedin.com/naman ◇ google-scholar ◇ github.com/McHarold404

EDUCATION

Arizona State University: Fulton School of Engineering — Tempe, Arizona
Master of Science in Computer Science, **GPA: 3.89/4** — Aug 2024 – Present
Research Area: *Table Generation, Semantic Parsing (NL2SQL), Agentic AI,Information Retrieval & Extraction*

International Institute of Information Technology, Hyderabad — Hyderabad, India
Bachelor of Technology (Honors) in Computer Science and Engineering, **GPA: 8.6/10** — 2019 – 2023
Research Area: *Applied NLP, Computational Social Science, Code-Mixing, Social Network Analysis*

Delhi Public School, Vasant Kunj — New Delhi, India
High-School Diploma, 12th Grade: 92.6% — 2005 – 2019
Gold Medal for Academic Excellence for 7 consecutive years

THESIS

Title: **Strategies for Structured Data Generation**
Transform unstructured text into reliable, auditable tables with a schema-first approach that couples structural planning with rule and time-aware reasoning over long contexts. Working on developing general frameworks, benchmarks, and training strategies for compact models to boost information coverage and faithful memory tracking, enabling trustworthy analytics and evidence synthesis across various domains.
Thesis Committee: **Dr. Vivek Gupta** (Committee Chair), **Dr. Chitta Baral**, **Dr. Huan Liu**

RESEARCH & PUBLICATIONS

Ahuja, N., Bardoliya, F., Baral, C., Gupta, V., *Map&Make: Schema-Guided Text-to-Table Generation*, ACL 2025 [Paper]

Martinez, S., **Ahuja, N.**, Bardoliya, F., Bryan, C., Gupta, V., *SportSQL: Real-Time Sports QA using NL2SQL over Dynamic Data*, AACL Demo Track 2025 [Paper]

Upadhyay, R., **Ahuja, N.**, Baral, R., Garimella, A., Gupta, V., *CMT-Bench: Cricket Multi-Table Generation Benchmark for Probing Robustness in Large Language Models*, Under Review (EACL 2025) [Paper]

Kodali, P., Bhatnagar, A., **Ahuja, N.**, Shrivastava, M., Kumaraguru, P., *HashSet: A Dataset for Hashtag Segmentation*, LREC 2022 [Paper]

Kodali, P., Sachan, T., Goindani, A., Goel, A., **Ahuja, N.**, Shrivastava, M., Kumaraguru, P., *Estimating Code-Mix Quality via LM Embeddings*, INLG 2022 (Generation Challenges Workshop) [Paper]

Sharma, S., Goyal, A., **Ahuja, N.**, Pati, B., Sharma, R., Kumaraguru, P., *From Noise to Knowledge: Distill Social Media Opinions on Rumor Detection Exploiting LLMs*, SSRN Working Paper, 2024 [Paper]

ACHIEVEMENTS

MITACS Research Scholar 2022: Selected for Globalink Summer Research Program to conduct research at Lakehead University, Canada.

Dean's Merit List 2021: Dean's Merit list for two consecutive semesters (Monsoon 2021, Spring 2022) for academic performance.

2nd place at Opportunity Hack 2024: Built a mentor-mentee scheduling system and integrated it with an NGO (Steam SuperHeroes).

JEE Mains: Achieved **All India Rank**(AIR)**1065** (99.91% percentile) in JEE Mains 2019.

JEE Advanced: Achieved **AIR 1578**, among 100,000 students in JEE Advanced 2019.

KVPY: Qualified for Kishore Vaigyanik Protsahan Yojana (KVPY) scholarship, 2017

NTSE: Qualified for National Talent Search Examination Scholarship, 2016.

EXPERIENCE

Amazon Web Services — May 2025 – Aug 2025
Software Development Engineer Intern (Manager: Samuel Nimako-Mensah) — Seattle, WA

- Designed a recommendation portal for developer tool discovery, reducing search latency from **1.5 minutes to 10s**.
- Built scalable backend services with Java + GraphQL + DynamoDB, accelerating developer onboarding for **100K+ engineers**.
- Optimized caching and load balancing, cutting page load time by **40%** and improving reliability across distributed teams.

CoRAL - Mayo Clinic Sep 2024 – May 2025
Research Assistant under Dr. Vivek Gupta (CoRAL), Dr. Irbaz Riaz (Mayo) Tempe, AZ

- Built an **agentic RAG system** for clinical trial PDFs with single-document insight extraction.
- Designed a novel prompting strategy for extracting and storing structured **table & image representations**.
- Developed a **structured-driven retrieval framework** for answer attribution and explainability.
- Benchmarked Agent using SOTA LLMs and achieved **+10% accuracy** vs OpenAI's file-search API at 1/3rd cost.

Perceptive Analytics Jun 2023 – Jun 2024
Analyst (Manager: Akhil A) Hyderabad, India

- Collaborated with a financial institution to build a **Retrieval-Augmented Generation (RAG) chatbot** for analyzing market trends on proprietary and SEC datasets.
- Built a **multi-agent LLM framework** with LangChain, automating insight extraction and market trend visualizations in Tableau.
- Designed an interactive interface with **Streamlit** and deployed as a serverless function on **AWS Lambda**.
- Reduced report turnaround time by **40%**, saving 25+ analyst hours weekly.

Subtl.ai Dec 2022 – Feb 2023
NLP Engineer (Manager: Vishnu Ramesh) Hyderabad, India

- Developed a **multi-user chatbot** for querying financial documents and Excel sheets, reducing manual analysis workload by **90%**.
- Deployed NLP pipelines with **Docker + Jenkins** and collaborated in **Agile sprints**, shipping incremental updates every 2 weeks.
- Improved query relevance and response time by **26%** through model fine-tuning (T5).

RELEVANT PROJECTS

Mentor-Mentee Platform [Github] — *Python, Flask, AWS (Lambda, DynamoDB, S3)* Oct 2024 – Feb 2025

- Developed dashboards for real-time performance tracking, driving actionable insights that increased engagement by **30%**.
- Designed an optimal mentor-mentee assignment system handling 6 different field constraints and 340 average weekly requests.
- Secured **2nd place** at Opportunity Hack 2024.

C-Shell (She Sells C Shells)[Github] — *C++* Sep 2022

- Implemented an advanced Linux shell in **C++** supporting I/O redirection, piping, and custom signal handling.
- Extended MIT's legacy **xv6** OS, adding priority-based and MLFQ scheduling.

Job Searching Web Application [Github] — *React.js, Node.js, Express, MongoDB* Jan 2022 – Mar 2022

- Built a full-stack portal enabling 200+ students to apply for jobs with a real-time **React dashboard**.
- Implemented fuzzy string matching to enhance job search accuracy and user experience.

Wikipedia Search Engine [Github] — *C++, Python* Jun 2021

- Engineered a multi-threaded indexer and search engine for **500K+ Wikipedia articles**.
- Implemented the **BM25 ranking algorithm**, achieving sub-**0.1s latency** for top-5 search results.

TEACHING AND PRESENTATIONS

Conference: at ACL 2025, presented my research paper Map&Make at the Information Extraction session.

Teaching Assistant, ASU: for *CSE 576: Topics in Natural Language Processing* (Spring 2025). Curated course assignments, exams, quizzes, and lecture slides. Mentored 20 MS students for course projects.

Teaching Assistant, IIIT-H: for *CS2.201: Machine Data and Learning* (Spring 2023). Oversaw recitations, labs, and course-project mentoring.

Guest Lecturer, ASU: for *CSE 576: Topics in Natural Language Processing* (Spring 2025). Delivered two lectures on Neural Networks and Backpropagation.