Mayank Goel

+91 789-054-5454 | themayankgoel28@gmail.com | linkedin | github

EDUCATION

International Institute of Information Technology - Hyderabad Hyderabad, India
B.Tech in Computer Science, CGPA: 7.96/10 *Jul. 2019 – Apr. 2023*

International Institute of Information Technology - Hyderabad Hyderabad, India
MS in Computational Linguistics *May 2023 – Dec. 2024*

EXPERIENCE

Researcher Aug. 2025 - Present
Lossfunk *Bangalore, India*

- Working on solving for moral realism through game theory and LLM simulations.

Research Engineer Mar. 2025 - Jul. 2025
IBM Research *Bangalore, India*

- Contributed to IBM's Granite LLM team, focusing on improving data quality for model training through advanced preprocessing techniques, synthetic data generation and ablation studies.
- Led efforts in distributed GPU training using FSDP, VLLM, and other model-sharding strategies to scale training and inference efficiently.

Researcher Jan. 2025 - Apr. 2025
AI Safety Camp *Remote*

- Part of a team selected by Prof. Masaharu Mizumoto of JAIST on a model to improve the morality of LLMs.
- Involved with discussions on philosophical grounding of research methodology, primarily around Virtue Ethics as a moral framework.
- Ran various finetuning experiments (using SFT/DPO) to test efficacy of our curated dataset.

Research Engineer Sep. 2024 – Mar. 2025
Deccan AI *Hyderabad, India*

- Involved with benchmarking of models on various tasks, across modalities, such as tasks on image, speech, code.
- Worked on automated QC for data generated by human annotators. Includes fine-tuning models on curated data.

Research Intern Oct. 2023 – Jul. 2024
ComplexData Lab, MILA AI *Remote*

- Designed experiments using the Debate framework for uncertainty quantification of LLM outputs.
- Worked on experiments to analyze and visualize data from other team projects on uncertainty quantification.
- Work done during this phase was accepted to COLM '24 and NeurIPS SafeGenAI Workshop '24

Teaching Assistant Jul. 2021 – Apr. 2024
IIIT Hyderabad *Hyderabad, India*

- Was involved with designing and evaluating assignments and conducting tutorials for the course Introduction to NLP in Spring '22 and Advanced NLP in Monsoon '23

Research Intern Jun. 2023 – Jul. 2023
Robert Bosch GmbH *Bangalore, India*

- Built upon an existing document hierarchal clustering project, add functionality to use sentence transformers.
- Built a robust pipeline to fine-tune sentence transformers using various methods.
- Developed a novel evaluation metric exclusively tailored to hierarchical clustering, addressing existing limitations with evaluation methods.

Research Associate Mar. 2021 – Feb. 2023
Proxzar AI *Remote*

- Fine-tuned Llama-7B on finance data using LORA and quantization. Did a hyperparameter search.
- Developed a robust e-commerce website crawling pipeline using Apache Nutch and Selenium
- Built a Spacy3 model utilizing Transformersfor a production-grade custom NER model for an e-commerce company, as part of an end-to-end system, with a dataset augmented by 20% using fuzzy matching. Achieved accuracy of 91% on a multi-classsystem

Research Intern   May 2021 – Jun. 2021
Trivedi Centre for Political Data, Ashoka University   *Remote*

- Built a streamlit dashboard to display visualized data of MPs and Governors in India, which accepts a variety of parameters. Work can be viewed at [link].

PUBLICATIONS

Conceptual Limitations of Current AI Safety Approaches and Virtue Ethics as an Alternative *Tokyo AI Safety Conference '25*
Authors: Masaharu Mizumoto, Rujuta Karekar, Mads Udengaard, **Mayank Goel**, Daan Henselmans, Nurshafira Noh, Saptadip Saha, Pranshul Bohra

Epistemic Integrity in Large Language Models *NeurIPS '24 SafeGenAI Workshop*
Authors: Bijean Ghafouri, Shahrad Mohammadzadeh, James Zhou, Pratheeksha Nair, Jacob-Junqi Tian, **Mayank Goel**, Reihaneh Rabbany, Jean-François Godbout, Kellin Pelrine

Web Retrieval Agents for Evidence-Based Misinformation Detection *COLM '24*
Authors: Jacob-Junqi Tian, Hao Yu, Yury Orlovskiy, Tyler Vergho, Mauricio Rivera, **Mayank Goel**, Zachary Yang, Jean-François Godbout, Reihaneh Rabbany, Kellin Pelrine

Joke Generation Using Masked Language Infilling on Automatically Extracted Templates *Humor Research Conference '23*
Authors: **Mayank Goel**, Abhijit Manatkar

Towards conversational humor analysis and design *Humor Research Conference '21*
Authors: Tanishq Chaudhary*, **Mayank Goel***, Radhika Mamidi

Automated Evalauation of Conversational Prompts using Ordinal Ranking for Humor Generation *ICON '24*
Authors: **Mayank Goel**, Parameswari Krishnamurthy, Radhika Mamidi

TECHNICAL SKILLS

Languages: Python (proficient), Bash, Javascript, C++
Frameworks: PyTorch, HuggingFace Transformers, Streamlit, Selenium, SpaCy

AWARDS AND OTHER RELEVANT INFORMATION

- **Research Award** from IIIT Hyderabad in 2021
- **Top 30** in the **Indian edition of International Olympiad of Linguistics**
- **Participant**, **ARENA 2024, Spring**
- **Participant**, **AI Safety Fundamentals Course, BlueDot Impact**
- **Ex-head** of University **Debate Society** and **Humor Club**
- **Founder-Moderator** of **Philosophy Friday Discussion Group**
- **Best Project, Theme Track** in **Natural Language API Hackathon by expert.ai**
- **1st Prize, Equity Track** in **HackUMBC, University of Maryland**
- **Best Web Application** in **HackAtHome, Brown University**
- **3rd Prize, Walmart Track** in **TAMU Datathon, Texas A&M University**
- **1st Prize, Stellantis Track** in **Megathon '22, IIIT Hyderabad**