

Manan Uppadhyay

+91 7650876355 | mananuppadhyay@gmail.com | [GitHub](#) | [Google Scholar](#) | [LinkedIn](#)

RESEARCH INTERESTS

I am a Research Fellow at **Microsoft Research India**, working with the multilingual NLP group. My research focuses on **cross-lingual parity in large language models**: understanding why model behavior differs across languages and tasks, measuring these gaps rigorously, and developing methods to close them. I am also deeply interested in using pragmatic interpretability findings to build robust, reliable and safe AI systems that benefit everyone across languages, cultures and regions.

EDUCATION

Birla Institute of Technology and Science, Pilani

B.E. in Electrical and Electronics Engineering (Minor in Data Science)

Goa, India

2020 – 2024

PUBLICATIONS & PREPRINTS

DEPART: DEcomposing PARiTy across Multilingual LLMs

EMNLP 2026

- Introduced a mixed-effects statistical framework to decompose and isolate the structural sources of cross-lingual performance disparities in multilingual LLMs.
- Demonstrated that representational similarity to English serves as a primary predictor of performance downstream, while tokenizer fertility yields highly task-dependent effects.
- Accepted at **EMNLP 2026 (Findings)**. Preprint: [arXiv:2605.28163](#).

The State and Fate of Multilingual, Contextual Evaluation in the NLP World

COLM 2026

- Conducted a rigorous data-driven audit of 51 multilingual benchmarks, revealing severe systemic biases: 36% of languages appear in only a single benchmark, and 56% of all evaluation data is translated from English.
- Quantified widespread data contamination across legacy evaluation suites, establishing clear guidelines for future benchmark design.
- Accepted at **COLM 2026**. Preprint: [Microsoft Research Link](#).

Samiksha

Technical Report

- A comprehensive, mixed-methods evaluation combining **150k native-speaker human annotations**, expert qualitative reviews, and **1.6 million automated LLM-as-a-judge evaluations**.
- Uncovered critical domain and language-specific performance trends across Indic models, exposing major divergences between automated metrics and human judgment.
- Technical Report: [Karya Evals Link](#).

UPDESH: Synthesizing Grounded IFT Data for 13 Indic Languages

ACL 2026

- Developed **UPDESH**, a massive 9.5 million-sample culturally grounded synthetic instruction-tuning dataset spanning 13 Indic languages, engineered to mitigate cross-lingual performance gaps.
- Accepted at **ACL 2026 (Main Track)**. Preprint: [arXiv:2509.21294](#).

Garuda: An Indian Agricultural LLM and Krishi Saathi Advisory Chatbot

AAAI 2025

- Designed and open-sourced an agricultural domain-specific LLM paired with a localized advisory chatbot framework for regional Indian farmers.
- Accepted as a demonstration paper at the Workshop on Open-Source AI for Mainstream Use at **AAAI 2025**.

COURSEWORK

Math: Probability and Statistics, Statistical Inference and Applications, Applied Stochastic processes, Applied Statistical Methods, Optimization, Linear Algebra

Machine/Deep Learning: Machine Learning, Foundations of Data Science, Large Language Models (In association with TCS Research), Digital Image Processing

EXPERIENCE

Research Fellow [Microsoft Research]

August 2025 – Present

- Language disparity in LLMs: Working on analysing and bridging the disparity across Languages for an LLM in multilingual settings. Work done includes an audit of **51 multilingual benchmarks** with a focus on coverage, representativeness and rigour
- DePART: Quantifying the language disparity in LLMs and framework for decomposing the sources of cross-lingual performance gaps in multilingual LLMs
- Multilingual Synthetic Data: Created a **9.5M** sample culturally grounded instruction tuning dataset, Updesh (8.6k downloads in the last month).
- Samiksha: A large-scale multilingual evaluation of Indic models combining **150k** native-speaker annotations and **1.6M** LLM-as-a-judge evaluations to surface divergences between automated metrics and human judgment.

Data Scientist [Razorpay]

Apr 2025 – Aug 2025

- Converted the existing payment routing system (**Doppler**) from classical ML to a real-time **multi-armed bandit reinforcement learning** framework to optimize transaction success rates.
- Designed and deployed **Airflow DAGs** for automated data retrieval and performance evaluation of routing systems.
- Conducted extensive **EDA** and rebuilt feature engineering pipelines to retrain classical models, improving daily payment success rate by **2%**.

Associate ML Scientist-1 [Wadhvani AI]

August 2024 – March 2025

- Developed a domain-specific language model for agriculture, utilizing fine-tuning techniques and synthetic data pipelines to improve performance, achieving a **1-5% improvement over GPT-4o-mini** on key internal benchmarks.
- Designed and deployed a chatbot for Kisan Call Centres (KCC), using an agentic LLM workflow to assist the advisors in quick query resolution for the farmers. The chatbot served **3,000 queries daily** during pilot runs across 11 states, with plans to scale further.
- Worked closely with the Kisan eMitra Chatbot, streamlining insurance related grievance resolution, covering **3 government schemes and supporting over 3 million queries in 8 months**, benefiting **250,000+ farmers**.

Machine Learning Intern [Wadhvani AI]

January 2024 – July 2024

- Implemented and compared Retrieval-Augmented Generation (RAG) and fine-tuning methods on models like Mistral-7B and LLaMA-8B for domain-specific downstream tasks.
- Designed and evaluated synthetic data pipelines for instruction tuning, enabling efficient model training.
- Evaluated and worked on improving LLM-generated responses in RAG-based setups to ensure high-quality performance and relevance.

Undergraduate Researcher [APPCAIR, UNSW]

August 2023 – January 2024

- Worked on building a 7 Million parameter deep learning model for detecting objects in endoscopic medical footage, deployable on edge devices.
- Implemented MobileOne architecture as the backbone with a detector on top, achieving **91% mAP@50** in bounding box detection on endoscopic datasets.
- Supervised by: Prof. Tanmay Verlekar, Prof. Ashwin Srinivasan, Prof. Eric Meijering, Prof. Manik Gupta.

PROJECTS AND INITIATIVES

Cross Tokenizer Distillation

March 2025

- Implemented the paper on Cross Tokenizer Distillation [ULD Loss] to enhance small-scale models by leveraging larger models as supervisors. (**Paper**)
- Trained bloomz-560m (student) using phi-3-mini-instruct (teacher) on the QED dataset, improving F1 score from **0.550 to 0.588**. (**GitHub code**)

An implementation of Einops

April 2025

- Implemented the einops library often used for tensor manipulations in plain numpy to support all rearrange and axis change operations (**The library**)
- The code and the sample test cases generated for the same could be found in the (**GitHub code**)

Object detection using Res-YOLO

- Applied the model architecture taken from the [paper](#), onto an open source car detection Dataset.
- Uses a ResNet backbone as a feature extractor, and uses a YOLO like architecture to detect cars.
- Fine-Tuned the architecture on a car detection dataset, by freezing the initial few ResNet layers and training the custom detector in a regression like fashion.
- Tech stack: Pytorch, numpy, matplotlib (**The code**)

COMMUNITY AND LEADERSHIP

Debate Head [The Literary and Debating Club BITS Goa]

- Was responsible for leading the debating wing of the college club and training the members to participate in multiple debating competitions.
- Participated in multiple National and international Debating competitions as a speaker and a judge
- Some Notable Achievements: Runners up at MSRMPD, and open semifinalist at IITB, IITK, BITS Pilani debate tournaments. Judged the elimination rounds at Imperial IV and Leiden Open 2022

TECHNICAL SKILLS

Languages: Python, C++, Java, C

Developer Tools: Git, Github, VS Code

Frameworks: Pytorch, TensorFlow, HuggingFace, Langchain, Llama Index, RAGAs

Libraries: numpy, pandas, scikit-learn, matplotlib, d3.js, dc.js

Applications: Autocad, Excel, Anaconda

Communication: English, Hindi