Debajyoti Mazumder

Indian Institute of Science Education and Research Bhopal, Bhopal Bypass Road, Bhauri, Bhopal 462 066, Madhya Pradesh, India

(+91) 7503393473 | debajyoti22@iiserb.ac.in | debajyotimaz@gmail.com | Website | LinkedIn | GitHub | Google Scholar

Summary

I am a final-year Ph.D. candidate in Data Science & Engineering at IISER Bhopal, working on multilingual and code-mixed natural language processing, representation learning, and the interpretability of large language models. My research examines how LLMs represent and reason over knowledge across languages and low-resource settings, with particular attention to cross-lingual consistency, robustness, and factuality. I combine analysis of model internals with applied methods such as fine-tuning, prompting, and reinforcement learning to study how models acquire, retain, and generalize knowledge. My work spans representation learning, interpretability, and automated fact-checking. My contributions include five first-author publications: four papers at EMNLP, a CORE A* conference, and one transaction (ACM TALLIP), along with papers at WSDM, AACL-IJCNLP, and The ACM Web Conference workshops. More broadly, I am interested in building NLP systems that are transparent, robust, and equitable across languages and cultures.

Research Interests

Multilingual & Code-Mixed NLP • Crosslingual and Crossmodal Knowledge Consistency • LLM Interpretability • Automated Fact-Checking • Representation Learning • Multi-Task Learning • Low-Resource NLP

Education

Indian Institute of Science Education and Research Bhopal (IISER Bhopal) Madhya Pradesh, India
Ph.D. in Data Science & Engineering; CPI: 9.00 / 10.00 *Aug. 2022 – Present*

- Supervisor: Dr. Akash Anil.
- Coursework: Machine Learning, Deep Learning, Natural Language Processing, Advanced Natural Language Processing, Spatial Data Science & Applications, Data Science in Practice.

Zakir Husain Delhi College, University of Delhi New Delhi, India
M.Sc. in Mathematics; Score: 69.25% *Jul. 2018 – Oct. 2020*

Deshbandhu College, University of Delhi New Delhi, India
B.Sc. (Honours) in Mathematics; SGPA: 8.70 / 10.00 *Jul. 2015 – Jul. 2018*

Kendriya Vidyalaya No. 2, Delhi Cantt. (CBSE Board) New Delhi, India
Senior Secondary; Score: 89.80% *2015*

Kendriya Vidyalaya No. 3, AFS-II, Jamnagar (CBSE Board) Gujarat, India
Secondary; CGPA: 9.60 / 10.00 *2013*

Research & Professional Experience

Adobe Research
Research Ph.D. Intern *Jul. 2026 – Sep. 2026*

- Built a next-question recommender for a conversational system for enterprise (CXO) users, whose backend is consumer-level aggregated tabular data.
- Used personalized PageRank for persona-based recommendation, providing context to the recommender system.
- Mentors: Iftikhar Ahamath Burhanuddin (Senior Research Scientist) and Atanu R Sinha (Principal Scientist).

NielsenIQ
Research Scientist – Senior Intern *Jul. 2025 – Jan. 2026*

- Worked on LLM reasoning using reinforcement learning with verifiable rewards (RLVR) for e-commerce tasks.
- Experimented with warm-up supervised fine-tuning and reasoning strategies, including variants of GRPO.
- Mentors: David Jiménez (Manager, Research Science) and Emilio J. Almazán (Lead Research Scientist).

IISER Bhopal Bhopal, Madhya Pradesh, India
Doctoral Researcher *Aug. 2022 – Present*

- Investigates how native-language signals improve understanding of Hindi-English code-mixed text.
- Develops methods for semantic tasks – humor, sarcasm, and hate-speech detection – in code-mixed settings.
- Studies LLM cross-lingual alignment, representation consistency, and interpretability.

Publications

- [**arXiv preprint 2026**] **D. Mazumder**, A. Anil. "A Two-Way Street: Surveying the Interplay between Code-Mixing and Native Languages in Text and Speech." [Paper]

  Organizes transfer work in both directions, native to code-mixed and the reverse, by direction, route, and modality. Finds that native data helps more than new architectures on a shared speech benchmark.

- [**AACL-IJCNLP 2026 Findings**] G. Kumar, A. Garg, **D. Mazumder**, A. Kishore, B. Kumar, J. Patro. "Entailed Opinion Matters: Improving the Fact-Checking Performance of Language Models by Relying on their Entailment Ability." [Paper]

  Leverages the entailment ability of language models for fact-checking; compares training-based and prompting-based methods.

- [**EMNLP 2026, Oral, Best Paper Nominee**] **D. Mazumder**, Mamta, A. Subramanyam. "To Trust or Not to Trust: Retrieval-Augmented Fact Checking in Speech." [Paper]

  Studies the transferability of the fact-verification ability of large audio-language models (LALMs) from text to speech.

- [**EMNLP 2026 Findings**] **D. Mazumder**, D. Pathak, P. Kodali, A. Joshi, A. Agarwal, J. Patro. "Evaluating Cross-lingual Knowledge Consistency in Code-Mixed vis-a-vis Indian Languages using IndicKLAR." [Paper][Project Page]

  Introduces IndicKLAR, an 18-language Indic benchmark for cross-lingual knowledge consistency, and Translate-in-Thought (TinT) prompting.

- [**EMNLP 2026 Findings**] **D. Mazumder**, D. Pathak, P. Kodali, J. Patro. "Neither Here Nor There: Cross-Lingual Representation Dynamics of Code-Mixed Text — A Hindi–English Case Study." [Paper][Project Page]

  Introduces a trilingual corpus to probe cross-lingual alignment, and a post-training alignment method motivated by interpretability analysis.

- [**DHOW-MiLLA @ TheWebConf 2026**] D. Gavit, G. Kumar, **D. Mazumder**, S. Das, J. Patro. "On VLMs for Diverse Tasks in Multimodal Meme Classification." [Paper]

  A comprehensive study of vision-language models (VLMs) via prompting and training approaches, with VLM-understanding-based LLM fine-tuning.

- [**ACM WSDM 2026**] D. Sharma, A. Kishore, A. Garg, **D. Mazumder**, D. Mohapatra, J. Patro. "Mind the Links: Cross-Layer Attention for Link Prediction in Multiplex Networks." [Paper]

  Introduces link prediction in the partial graph of multiplex networks as a novel task.

- [**EMNLP 2025 Findings**] **D. Mazumder**, A. Kumar, J. Patro. "Revealing the Impact of Synthetic Native Samples and Multi-Tasking Strategies in Hindi-English Code-Mixed Humour and Sarcasm Detection." [Paper]

  A multi-task architecture that shares knowledge across related tasks and outperforms current state-of-the-art models.

- [**ACM TALLIP 2025**] **D. Mazumder**, A. Kumar, J. Patro. "Improving Code-Mixed Hate Detection by Native Sample Mixing: A Case Study for Hindi-English Code-Mixed Scenario." [Paper]

  Adds native-language hate samples to the training set, and transfers knowledge from native-language samples alone to detect code-mixed hate.

Technical Skills

- **Methods**: Prompting, probing, fine-tuning, multi-task learning, reinforcement learning, personalized PageRank, LLM interpretability
- **Applications**: Humor, sarcasm, and hate speech detection, factual recall, fact checking, meme classification, link prediction, crosslingual alignment
- **Tools**: Python, PyTorch, Hugging Face, LaTeX, Git
- **Languages**: English, Hindi, Bengali

Honors & Awards

- **Best Paper Award Nominee**: EMNLP 2026, for "To Trust or Not to Trust" (VeriSpeak), Budapest, Hungary.
- **Student Volunteer and Diversity & Inclusion**: EMNLP 2026, Budapest, Hungary.
- **ACM/IARCS Travel Grant**: EMNLP 2026, Budapest, Hungary.
- **Selected, Machine Learning Summer School (MLSS)**: Columbia University, 2026, New York, USA.
- **SIGIR Student Travel Grant**: WSDM 2026, Boise, Idaho, USA.
- **ACM Travel Grant (Invited Talk & Poster Presenter)**: ACM ARCS 2026, IIT Hyderabad.
- **ANRF ITS Travel Grant (Poster Presenter)**: EMNLP 2025, Suzhou, China.
- **Selected, LiveRAG Challenge Track**: SIGIR 2025, Padua, Italy.

- **Travel Grant (Participant)**: Indian Symposium on Machine Learning (IndoML 2023), IIT Bombay.
- **2nd Position, Poster Presentation**: DSE Day 2023, IISER Bhopal, Bhopal, Madhya Pradesh, India.
- **Travel Grant (Participant)**: Indian Symposium on Machine Learning (IndoML 2022), IIT Gandhinagar.
- **MHRD Fellowship Recipient**: GATE 2022, India.

Teaching and Academic Service

IISER Bhopal — Bhopal, Madhya Pradesh, India
Teaching Assistant — *Jan. 2023 – Nov. 2026*

- **Machine Learning with Graphs** (DSE 423/623): Aug.–Nov. 2026.
- **Natural Language Processing** (DSE 318/401/607): Jan.–Apr. 2026; (DSE 407/607): Aug.–Nov. 2024, Aug.–Nov. 2023.
- **Introduction to Programming** (ECS 102): Jan.–Apr. 2025, Jan.–Apr. 2024.
- **Applied Optimization** (DSE 311): Jan.–Apr. 2023.

Reviewing

- *Conferences:* ACL Rolling Review (ARR) Oct. 2026, ICWSM 2026, The ACM Web Conference 2026 and 2027, INSTCon 2026.
- *Workshops:* LT-EDI 2026, DravidianLangTech 2026.
- *Journals and Transactions:* IEEE Transactions on Multimedia.
- *Ethics Review:* ACL Rolling Review (ARR) Oct. 2026.

References

- **Dr. Akash Anil**, Ph.D. supervisor, Assistant Professor, Department of Data Science and Engineering, IISER Bhopal, India. anila@iiserb.ac.in
- **Dr. Jasabanta Patro**, Visiting Professor, School of Engineering and Technology, Rajiv Gandhi National Aviation University, Amethi, India. jasabantapatro@gmail.com
- **Dr. Aditya Joshi**, Senior Lecturer, School of Computer Science and Engineering, UNSW Sydney, Australia. aditya.joshi@unsw.edu.au