Paramita Das, Ph.D.

https://paramita08.github.io/portfolio/ dasparamita1708@gmail.com +91-9038877509

Paramita Das paramita08

AI Researcher and Computational Data Scientist with 6+ years of experience in Natural Language Processing, Multi-modal AI, and scalable Machine Learning solutions. Skilled in deploying Large Language Models (LLMs) and Vision Language Models (VLMs) on web corpora of collaborative platforms such as Wikipedia and Wikidata, ensuring a balance between quality and quantity of the platforms. Experienced in publishing at top-tier conferences, with expertise in high-dimensional data analysis, predictive modeling, and bridging research with real-world applications. Passionate about responsible AI and committed to translating research into impactful digital solutions.

Skills

Research Interest: Machine Learning, Deep Learning, Natural Language Processing, Generative AI, Responsible AI, Algorithms.

LLM & VLM Models: GPT, LLaMA, Mistral, CLIP, LLaVA.

Frameworks & Tools: LangChain, ChromaDB, PyTorch, TensorFlow, Jupyter, REST APIs.

Database & Cloud: AWS, Google Cloud, Hugging Face Hub.

Programming: Python, SQL, Java, PySpark.

Education

Indian Institute of Technology Kharagpur *Jan 2019 — March 2025*
Ph.D. (*Computer Science and Engineering*)

- **Thesis title**– Collaborative Challenges in Wikipedia and Wikidata: Striking a Balance Between Quality and Quantity
- **Supervisor:** Prof. Animesh Mukherjee

Indian Institute of Engineering Science & Technology Shibpur *Aug 2016 – Jun 2018*
M.Tech in Computer Science and Engineering

- **CGPA:** 8.26 out of 10.0
- **Dissertation topic:** Computational Data Science

Maulana Abul Kalam Azad University of Technology, West Bengal *Aug 2011 – Jun 2015*
B.Tech in Computer Science and Engineering

- **CGPA:** 9.05 out of 10.0

Higher Secondary (12th) Examination *2011*

- **GPA:** 88.4%

Work Experience

Research Associate *Sept 2024 - Present*
Indian Institute of Technology Kharagpur

- Working on addressing key challenges in detecting and reducing social biases in SOTA Vision-Language Models (VLMs), such as CLIP, LLaVA.
- Dataset used: Wikipedia images of popular persons.

Research Intern *Apr 2022 – Jun 2022*
Wikimedia Foundation

- Proposed a language-agnostic predictive model for assessing the quality of Wikipedia articles across multiple languages.
- Created a large dataset of language-agnostic features and predicted quality scores for approximately 2 billion article revisions across over 300 language editions of Wikipedia, covering revisions up to the end of 2022,

using the proposed model.

Project Linked Person *Sept 2018 - Dec 2018*
Indian Statistical Institute, Kolkata
- Worked on addressing research challenges in Image Processing, Remote Sensing.

Publications

- **Social Biases in Knowledge Representations of Wikidata separates Global North from Global South**: ***Paramita Das***, Sai Keerthana Karnam, Aditya Soni, Animesh Mukherjee, ***WebSci 2025***. (Core B)
- **On the effective transfer of knowledge from English to Hindi Wikipedia**: ***Paramita Das***, Amartya Roy, Ritabrata Chakraborty, Animesh Mukherjee, ***COLING 2025***. (Core B)
- **Language-Agnostic Modeling of Wikipedia Articles for Content Quality Assessment across Languages**: ***Paramita Das***, Isaac Johnson, Diego Saez-Trumper, Pablo Aragón, ***ICWSM 2024***. (Core A)
- **Diversity matters: Robustness of bias measurements in Wikidata**: ***Paramita Das***, Sai Keerthana Karnam, Anirban Panda, Bhanu Prakash Reddy Guda, Soumya Sarkar, Animesh Mukherjee, ***WebSci 2023***. (Core B)
- **Quality Change: Norm or Exception? Measurement, Analysis and Detection of Quality Change in Wikipedia**: ***Paramita Das***, Bhanu Prakash Reddy Guda, Sasi Bhusan Seelaboyina, Soumya Sarkar, Animesh Mukherjee, ***CSCW 2022***. (Core A)
- **When expertise gone missing: Uncovering the loss of prolific contributors in Wikipedia**: ***Paramita Das***, Bhanu Prakash Reddy Guda, Debajit Chakraborty, Soumya Sarkar, Animesh Mukherjee, ***ICADL 2021***. (Core A)
- **Mining the online infosphere: A survey**: S Adak, S Chakraborty, ***Paramita Das***, M Das, A Dash, R Hazra, B Mathew, P Saha, S Sarkar, A Mukherjee, ***Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 2022***. (Impact Factor- 10.30)
- **REVerSum: A Multi-staged Retrieval-Augmented Generation Method to Enhance Wikipedia Tail Biographies through Personal Narratives**: Sayantan Adak, Pauras Mangesh Meher, ***Paramita Das***, Animesh Mukherjee, ***COLING 2025***. (Core B)
- **Accessibility metric for characterizing the relevance of conference papers**: ***Paramita Das***, Agniv Adhikari, Abhik Mukherjee, ***IEEE Region 10 Symposium (TENSYMP) 2019***.
- **Generating a representative keyword subset pertaining to an academic conference series**: Agniv Adhikari, ***Paramita Das***, Abhik Mukherjee, ***Scientometrics 2019***. (Impact Factor- 3.8)

Academic Services

Teaching Assistant: Social Computing, Information Retrieval, Complex Networks, Machine Learning, Theory of Computation, Theory of Algorithms.

Reviewer: ACM Multimedia, CSCW, EACL, IEEE INDICON, TKDE, Wiki Workshop.

Achievement

- Received research fellowship from the Ministry of Education, India, as financial assistance for pursuing the Ph.D.
- Qualified UGC-NET for Assistant Professor in 2018.
- Awarded scholarship by the Government of India for outstanding performance in the Higher Secondary Examination (2011).
- Elected member of the Student Senate at IIT Kharagpur for 2020 and 2021.

References

Prof. Animesh Mukherjee
Professor
Department of CSE, IIT Kharagpur
animeshm@cse.iitkgp.ac.in

Diego Sáez Trumper
Senior Research Scientist
Wikimedia Foundation
diego@wikimedia.org