Himanshu Dutta

himanshu-dutta.github.io | himanshudutta | himanshudutta

Research Fellow | Microsoft Research

hdutta1024@gmail.com | Himanshu Dutta

Education

- **Indian Institute of Technology, Bombay**
  MTech, Computer Science and Engineering *GPA: 9.69/10* Aug'23 – June'25
  Advisors: Prof. Pushpak Bhattacharyya & Prof. Preethi Jyothi
- **KIIT University, Bhubaneswar**
  BTech, Computer Science & System Engineering *GPA: 9.65/10* (**Department Rank 2**) Aug'18 – June'22

Publications & Preprints

(* = equal contribution)

- *GRAFT: Graph-based Flow-aware Document-level Machine Translation* **EMNLP (Industry) '25**
  Himanshu Dutta, Sunny Manchanda, Pushpak Bhattacharyya, *et al.* paper code
- *Recon, Answer, Verify: Agents in Search of Truth* **EMNLP (Industry) '25**
  Satyam Shukla*, **Himanshu Dutta***, Pushpak Bhattacharyya paper
- *Genetically Optimized UFLANN for Uncovering Clusters* **IEEE Access '23**
  Himanshu Dutta*, Satchidananda Dehuri, Ashish Ghosh, *et al.* paper
- *Feature Selection by Black Widow Optimization for Classification* **ICCIPR '22**
  Himanshu Dutta*, Mahendra Kumar Gourisaria, Himansu Das paper code
- *EnsembleDet: Ensembling Against Adversarial Attack on Deepfake Detection* **JEI '21**
  Himanshu Dutta*, Aditya Pandey*, Saurabh Bilgaiyan paper

Research Experience

Research Fellow | Microsoft Research July'25 – Present
Hosts: Dr. Lokesh Nagalapatti, Dr. Yashoteja Prabhu, & Dr. Manik Varma

- Developed novel *Efficient Negative Sampling and Batch Sampling* algorithms to train large-scale dense retrieval models.
- With our developed batch sampling algorithm, we demonstrate a *gain of* $\sim$ **4** *NDCG@10 points on retrieval benchmarks.*
- Working with **ReFoRM** (retrieval) team to apply the developed techniques to Microsoft Bing Ads and Search products.

AI/ML Research Intern | Defence Research and Development Organisation June'24 – July'24
Host: Sunny Manchanda

- Developed LLM Agents and Multi-agent frameworks for Document-Level Machine Translation.
- Experiments for domain and country adaptation of MT, showed an improvement of **7.31** and **10.02** BLEU scores.
- Demonstrated efficacy of Agentic approaches by conducting a human preference evaluation study, showing Agentic setup is preferred **25%** of the time, as opposed to reference translation being preferred only 18% of the time.

Research Projects

Retrieval Augmented Generation & Machine Translation Jan'24 – June'24
Advisors: Prof. Pushpak Bhattacharyya & Prof. Preethi Jyothi

- Developed RAGMT, a multi-task Retrieval Augmented Generation (RAG) framework for Machine Translation (MT).
- Demonstrated its effectiveness in domain-adaptation of MT, achieving an *average improvement of 12.90 BLEU scores.*

Knowledge Infused Neural Machine Translation Oct'23 – Jan'24
Advisors: Prof. Pushpak Bhattacharyya & Prof. Preethi Jyothi

- Explored Translation Memory, Dictionary-based, and RAG approaches for knowledge infusion in Machine Translation.
- Implemented and customized *kNN-MT* for translation tasks using RAG.
- Conducted a comparative study, showing a **5.01** BLEU score improvement over Translation Memory-based.

Technical Skills

Languages	C, Python, C++, Bash, JavaScript, TypeScript
Tools & Frameworks	Pytorch, Lightning, Huggingface, LangChain, FairSeq2, FastAI, Tensorflow, Streamlit, Django, React JS, React Native

Scholastic Achievements

- Achieved **All India Rank 16** in **GATE Computer Science 2023** amongst 75,680 candidates.
- Received **KIIT Merit Scholarship** for achieving 2^{nd} **rank** in the batch for four semesters and 3^{nd} **rank** in one semester.