Swaroop Nath swaroop-nath.github.io
Pre-Doctoral Researcher | Google DeepMind
swaroopnath swaroopnath
swaroopnath6@gmail.com Swaroop Nath

Education

- **Indian Institute of Technology, Bombay**
  MS by Research, Computer Science and Engineering *GPA: 9.76/10* (**Department Rank 2**) Aug'21 – June'24
  Advisors: Prof. Pushpak Bhattacharyya & Prof. Harshad Khadilkar
  Recipient of the "**Winifred B. Fernandes Research Excellence Award**" for my MS Thesis.

Publications & Preprints

(* = equal contribution)

- One Prompt To Rule Them All: LLMs for Opinion Summary Evaluation **ACL'24**
  Swaroop Nath*, Tejpalsingh Siledar*, Pushpak Bhattacharyya, *et al.* paper code
- Replaces Supervision: Query focused Summarization using Deep Reinforcement Learning **EMNLP'23**
  Swaroop Nath, Pushpak Bhattacharyya, and Harshad Khadilkar paper code slides
- Transformers are Expressive, But Are They Expressive Enough for Regression?
  Swaroop Nath, Harshad Khadilkar, and Pushpak Bhattacharyya arxiv code
- Leveraging Domain Knowledge for Efficient Reward Modelling in RLHF: A Case-Study in E-Commerce Opinion Summarization.
  Swaroop Nath*, Tejpalsingh Siledar, Pushpak Bhattacharyya, *et al.* arxiv code slides

Research Experience

Pre-Doctoral Researcher | Google DeepMind July'24 – Present
Hosts: Dr. Praneeth Netrapalli, Dr. Nitish Gupta, & Dr. Prateek Jain

- Implemented a novel training-free, clustering-based *Efficient Attention algorithm* that led to $\sim$ **40**% reduction in latency while maintaining text generation quality.
- Developed a novel *Efficient Reasoning algorithm* which reduces verbosity in frontier reasoning models by $\sim$ **50**% while maintaining reasoning quality.

AI/ML Research Intern | LinkedIn May'23 – July'23
Host: Smit Marvaniya

- Developed a *novel* algorithm for Spam detection in LinkedIn comments, identifying the relevance of context (neighboring comments and post), which led to $\sim$ **20**% quality improvement of the overall pipeline.

Scholastic Achievements & News

- Invited for a talk at **Hyperbots** on my MS Thesis. slides July'24
- **Best talk** award for "*Efficient Reward Modelling in RLHF*" at RISC24, CSE@IIT-Bombay. slides Jan'24
- Delivered part of a **Tutorial on Vision-Language Models** at *International Conference on Natural Language Processing 2023* (*ICON 2023*). Dec'23
- Delivered a tutorial at **Qualcomm** on *RLHF* and *Alignment Research*. slides Aug'23
- Led IIT Bombay teams to **Silver and Bronze Medals**, in Inter IIT Tech Meet (among 22 IITs) Dec'22 & Jan'23
- Achieved a **perfect 10/10** GPA in the 2^{nd} semester (Jan'22 – May'22), with 5 graduate level credit courses.

Research Projects

Efficient Reward Modelling for RLHF by Leveraging Domain Knowledge Oct'23 – June'24
Advisors: Prof. Pushpak Bhattacharyya & Prof. Harshad Khadilkar

- Developed a novel Reward Modelling approach for Reinforcement Learning from Human Feedback, which reduced preference data requirement by **20**×, while still achieving alignment with humans when used in RLHF.
- Released two datasets for Opinion Summarization: PROMPTOPINSUMM – SFT data, & OPINPREF – Preference data.

Can Transformers approximate Smooth Functions? Oct'23 – Jan'24
Advisors: Prof. Pushpak Bhattacharyya & Prof. Harshad Khadilkar

- Investigated the expressivity of Transformers for Smooth Functions. Verified *theoretically* and *empirically* that Transformers cannot practically approximate smooth functions with finite layers.

Hierarchical Text Generation based Query focused Summarization (QfS) June'23 – Sept'24
Advisors: Prof. Pushpak Bhattacharyya & Prof. Harshad Khadilkar

- Experimented with Hierarchical Text Generation for QfS, motivated by the underlying textual structure present within summaries.

Reinforcement Learning based Query focused Summarization (QfS) May'22 – June'23
Advisors: Prof. Pushpak Bhattacharyya & Prof. Harshad Khadilkar

- Developed a novel Reinforcement Learning algorithm for QfS – led to **37.2**% improvement (automatic metric) and **19**% improvement (human evaluation).
- Contributed a *gold-standard* dataset (**250 samples**) which tackles Topic Centralization, for analysis of QfS models.

Passage Embedding using Siamese BERT Oct'22 – June'23
Advisors: Prof. Pushpak Bhattacharyya & Prof. Harshad Khadilkar

- Developed a novel *Passage Embedding mechanism* inspired by *Cluster Hypothesis*, which was used in a reward function for RL based QfS training – led to **2.21** point improvement (ROUGE) over the next best reward function.
- Contributed a *new* dataset ($\sim$ **8 million instances**) for pre-training the Passage Embedding model.

Open Domain Aspect-Based Sentiment Analysis (ABSA) Aug'21 – May'22
Advisor: Prof. Pushpak Bhattacharyya

- Created a new dataset for ABSA from Yelp reviews, motivated by the lack of diversity of domains in existing datasets. It covered **111 domains**, which led to $\sim$ **18**% improvement for models in *open-domain* performance.

Selected Course Projects

Beyond BranchFormer: What re-stitching the network does! Jan'23 – May'23
Instructor: Prof. Preethi Jyothi

- Patched the *faulty* dropout implementation in the original BranchFormer implementation, and experimented with both *Attention dropout* and *CG-MLP dropout*.
- Tested Linear Attention (from BigBird), achieving a CER of **4.97** (vs **4.22** in the original FastFormer Attention).

Caption Consistency Detection Jan'22 – May'22
Instructor: Prof. Pushpak Bhattacharyya

- Developed a CNN+Transformer-based model for the *novel* task: **Caption-Image Consistency Detection**.
- Implemented a custom **near-miss-near-hit** based sampling technique to mine negative samples for training, which helped the model achieve an F1-score of **89**%.

DoodlerGAN: Exploring GANs for Doodling July'21 – Nov'21
Instructor: Prof. Preethi Jyothi

- Reimplemented DoodlerGAN, a StyleGAN2 inspired recurrent part generator, for the Creative Birds dataset, which led to **75**% more diverse generations than the vanilla StyleGAN2 model.
- Explored the impact of ProGAN style progressive generation on the *convergence time* and *sample efficiency*.

Service

Reviewer for A/A* conferences Feb'24 – Present

- Served as a reviewer for several A/A* CL conferences (ACL, EMNLP, EACL, etc.), and a co-reviewer in NeurIPS'24.
- Nominated as a **great reviewer** for 3 reviewed papers (out of 4) for ACL'25.

Web Chair | CODS-COMAD Conference Sept'22 – Jan'23

- Coordinated with the organizing team to keep the conference website up-to-date for all information to be disseminated to the authors; and smooth running of the conference proceedings.

System Administrator | DGX Servers @ IoE-DIS, IIT Bombay Sept'22 – Dec'23

- Served as the **system administrator**, enabling smooth operation of the DGX-A100 servers at IIT Bombay.

Public Relations Manager | CFILT Lab, IIT Bombay Feb'22 – June'24

- Tasked with handling the CFILT Social Media outlets (LinkedIn, Twitter and Instagram). Increased online visibility by **619**% on LinkedIn over the duration.

Lead Teaching Assistant | Advanced-NLP | Qualcomm Apr'23 – Sept'23

- Helped in delivering a course on NLP to Qualcomm Engineers. Tasks handled: Preparing the Curriculum, Coding Demos and Preparing Quizzes.

Technical Skills

Languages	C, Java, Python, C++, Bash, Javascript
Tools & Frameworks	PyTorch, JAX, Flax, TensorFlow, HuggingFace, Ray, DeepSpeed, Streamlit, Gradio, Spring