Arnav Goel

+1-(412)-660-2046 | arnavgoe@cs.cmu.edu | Google Scholar | arnav10goel.github.io | linkedin.com/in/arnavg10 | Github

Education

Carnegie Mellon University — Pittsburgh, PA
Master of Science in Machine Learning, Machine Learning Department (SCS) — August 2025 - December 2026

- Coursework: Advanced Machine Learning, Deep Learning Systems (in progress)

Indraprastha Institute of Information Technology (IIIT) Delhi — New Delhi, India
B.Tech in Computer Science and Artificial Intelligence — September 2021 - June 2025

- CGPA: 9.50/10.00 **(Department Rank: 2, Institute Rank: 3)**
- Dean's List Award for Academic Excellence (all semesters); **Teaching Assistant:** Machine Learning (PG)
- Coursework: Data Structures, Database Management Systems, Object-Oriented Programming, Algorithm Design and Analysis, Operating Systems, Data Science, Convex Optimization, Machine Learning, Natural Language Processing, Information Retrieval, Deep Learning, Large Language Models, Reinforcement Learning

Publications (All first/co-first author)

1. **Optimizing Vision Language Models for Scientific VQA through Caption-Aware Supervised Training.** *AI4EDU Workshop @ AAAI-25*
2. **Attributing Culture-Conditioned Generations to Pretraining Corpora**. *Accepted to ICLR 2025* | [pdf] | [code]
3. **Exploring Multilingual Unseen Speaker Emotion Recognition: Leveraging Co-Attention Cues in Multitask Learning.** *Accepted to INTERSPEECH 2024* | [pdf] | [code]
4. **Multilingual Prosody Transfer: Comparing Supervised & Transfer Learning.** *ICLR 2024 (Tiny Track) - Invite to Present* | [pdf] | [code]
5. **CrossVoice: Crosslingual Prosody Preserving Cascade-S2ST using Transfer Learning.** *ICLR 2024 (Tiny Track) - Invite to Present* | [pdf] | [code]
6. **LLMGuard: Guarding Against Unsafe LLM Behavior.** *AAAI 2024 (Demo Track)* | [pdf]
7. **SciPhyRAG - Retrieval Augmentation to Improve LLMs on Physics Q&A.** *11th International Conference on Big Data and Artificial Intelligence (BDA 2023)* | [pdf] | [code]
8. **Revolutionizing High School Physics Education: A Novel Dataset.** *11th International Conference on Big Data and Artificial Intelligence (BDA 2023)* | [pdf]

Work Experience

Microsoft Research — January 2025 - July 2025
Research Engineering Intern | Supervisor: Dr. Amit Sharma, Dr. Manik Varma (Extreme Classification Group) — Bangalore, India

- Built **HORIZON**, a large-scale user modeling benchmark from long-range cross-domain histories, with optimized big-data handling (**Parquet**, **SQLite**). Proposed a novel **FAISS**-accelerated dual-encoder retrieval setup to evaluate **LLM-based recommendation / user-modeling systems**.
- Devised a **multi-step chain-of-thought** framework with **GPT-4o** for Bing Ad recommendations, improving **interestingness by 15%** and **relevance by 5%**. Benchmarked SLMs and reasoning models on HORIZON, scaling inference using **distributed** multi-GPU setups **(vLLM)**. Work under review at **NeurIPS 2025**.
- Led migration of **Bing Ads** recommendations from GPT-4o to lightweight **8B models**, designing a **synthetic reasoning chain generation** and **task-specific distillation** pipeline, achieving a **100× inference speedup** with teacher-level recall and precision.

University of Southern California — May 2024 - August 2024
IUSSTF-Viterbi Research Intern | Supervisor: Dr. Xiang Ren — Los Angeles, CA

- Designed MEMOED, the first framework attributing LLM cultural knowledge to **memorization vs. generalization** using **distribution analysis, signal-noise metrics, and IR techniques**, revealing how pre-training data distributions influence memorization and enabling **corpus quality and bias assessment**. First-author work accepted at **ICLR 2025**.
- Led the first large-scale analysis of LLM pre-training corpora by scaling MEMOED to billions of documents via **sliding-window time optimizations**, custom batching for API calls, and parallelized generation-evaluation pipelines, achieving a **50× speedup**.

IBM Research — May 2023 - October 2023
Research Engineering Intern | Supervisors: Dr. Sameep Mehta, Dr. Nishtha Madaan — New Delhi, India

- Developed lightweight LLM guardrails for enterprise applications, implementing modules for **blacklisting user-suggested topics** (87% accuracy via LDA, and keyword extraction), **toxicity detection** (bidirectional encoder achieving 91–95% accuracy on benchmark datasets), and **RAG-based hallucination detection** using Pinecone.
- Designed and deployed a **React/JS web-GUI** for enterprise-safe LLM interactions with **REST-based API** implementations for the guardrails, integrating **red-teaming**, and evaluation workflows for factual and domain-grounding. Accepted for demo at **AAAI-2024.**

MIRAE AI Systems Pvt. Ltd. — November 2023 - May 2024
AI/SWE Intern — New Delhi, India

- Designed efficient speech-to-speech translation pipelines using **OpenAI's Whisper with Flash Attention**; developed a novel gender bias-reduction method (**patent pending**) in translation systems via data and alignment-based interventions.
- Curated a morphology-aware evaluation benchmark to assess translation of grammatically gendered languages in LLMs.

Academic Research Experience

FORUM Lab, Carnegie Mellon University July 2025 - present
Graduate Research Assistant | Supervisor: Dr. Aditi Raghunathan Pittsburgh, PA

- Working on investigating **emergent misalignment** in instruction-tuned LLMs and data distributions aggravating it.

BLENDER Lab, University of Illinois at Urbana-Champaign June 2025 - present
Research Intern | Supervisor: Dr. Heng Ji Urbana, IL (Remote)

- Working on optimizing and evaluating **energy-based transformers** and **diffusion LMs** for self-verifiable reasoning.

Mila-Quebec Artificial Intelligence Institute July 2024 - January 2025
MITACS Research Intern | Supervisor : Dr. Jackie Chi Kit Cheung Montreal, Canada (Remote)

- Selected for the **MITACS Globalink Research Internship** at MILA, Quebec, focusing on erasure in natural language generation systems. Work under review in ACL-ARR.
- Worked in collaboration with **Microsoft Research Montreal's FATE team** to develop a qualitative taxonomy for categorizing erasure, contributing to enhanced evaluation of harm and reliability in NLP systems.

Infosys Center for Artificial Intelligence, IIIT-Delhi May 2023 - Jul 2024
Undergraduate Researcher | Supervisor: Dr. Anubha Gupta New Delhi, India

- Designed a prosody-preserving speech-to-speech translation system using Whisper-Fast, Meta's MMS-TTS, and X-vector voice cloning, achieving **15% higher MOS** and **+8 BLEU** over direct-S2ST systems. Published at **ICLR 2024's Tiny Track**.
- Proposed a **co-attention multi-task learning framework** in PyTorch for speaker emotion recognition, improving **F1 scores by 8%** on unseen multilingual speaker profiles over SOTA models. Published at **INTERSPEECH 2024**.

Multimodal Digital Media Analysis Lab (MIDAS-IIITD) January 2023 - May 2024
Undergraduate Researcher | Supervisor: Dr. Rajiv Ratn Shah New Delhi, India

- Enhanced scientific QA and reasoning in open-source LLMs via retrieval-augmented fine-tuning with curated corpora, improving final answer accuracy and BERTScore by **16 points** indicating higher reasoning quality. Published at **BDA 2023**
- Proposed a novel image-text alignment strategy using image descriptions to improve multimodal representations in open-source multimodal LLMs (VLMs), boosting final answer accuracy and reasoning performance by **7%** on complex Q&A benchmarks like **MMLU** and **SciQA**. Accepted at the **AI4Edu Workshop at AAAI-2025**.
- Led the development of (1) the first Indian textbooks-based knowledge corpus for physics using OCR techniques and (2) text and image-based Hindi and English scientific question answering datasets for evaluating reasoning performance.

Selected Projects

Amazon ML Challenge 2024 - Multimodal LLM OCR October 2024 - November 2024
Tools: Multimodal LLMs, Large-Scale Data Handling, Ensemble Learning Presentation |[]

- Secured a **Top-10 finish (6th rank in India)** in the **Amazon ML Challenge 2024** out of **20,000+ teams**, earning a presentation opportunity with Amazon scientists. Built a novel **few-shot learning** pipeline using **ensembled open-source VLMs** for entity extraction from 1.2M e-commerce images, achieving a **0.72 F1-score**.

Cross Lingual Transfer Learning for Indian Languages February 2024 - May 2024
Tools: Multilinguality, Pytorch, NLP []

- Investigated **cross-lingual transfer learning** for Indian languages, designing a novel **cross-attention module** in PyTorch to boost transfer in low-resource settings. Achieved **10–12% improvement** across three downstream cross-lingual tasks.

Multimodal Browser Agent February 2024 - May 2024
Tools: LLMs, Langchain, Pinecone, Unstructured Data Handling []

- JavaScript & CSS-based Chrome extension with real-time web scraping, **Pinecone**-based **multi-index retrieval**, and **multimodal LLM-based RAG** for context-aware, **multimodal Q&A**. Enabled user personalization via controlled browser history exposure and multi-tab support through caching using **LlamaIndex**.

Tank Stars: 2 Player Game August 2022 - December 2022
Tools: JAVA, Object-Oriented Programming, Design Patterns []

- Developed a 2-player game with **JAVA** (LibGDX), applying OOPs principles, design patterns *(Facade, Singleton, Flyweight)*, GUI, sound effects, and serializable game states, with **JUnit-testing** and UML diagrams.

Achievements and Other Experience

- **Google DeepMind Research Symposium 2025**: Selected among top 150 undergraduates in India.
- **Amazon ML Challenge 2024:** Ranked **6th** out of 20,000+ teams nationwide in India, securing an Applied Science Internship.
- **IUSSTF-Viterbi India Program 2024:** Selected among **15 nationwide scholars** for fully-funded research internship at University of Southern California's NLP group (INK Lab) due to prior research work in NLP and LLMs.
- **MITACS Globalink Research Scholarship 2024:** Awarded fully-funded research internship at **MILA-Quebec AI Institute** and **McGill University** for prior research in NLP.
- **Dean's List for Distinguished Academic Excellence 2024:** Awarded for being the **top ranker (A+)** in 3 courses *(NLP, Information Retrieval, Data Science)* in one academic year. Only awardee from my batch.
- **Amazon ML Summer School 2023:** Selected for the prestigious, annual summer school conducted by Amazon.
- **Summer Undergraduate Research Fellowship 2023:** Honored by IRD-IIIT Delhi for my research project on improving scientific question answering abilities of LLMs *(SciPhyRAG)*.
- **Joint Entrance Examination**: JEE Main: **99.87** percentile; JEE Advanced: **99.76** percentile out of 1.1 million candidates
- **Served as Reviewer**: ACM Multimedia 2024, ICLR 2024 (Tiny Track), COLING 2025, ACL 2025, EMNLP 2025

Technical Skills

Languages: Python (Numpy, Pandas, Matplotlib, Scikit-Learn, Scipy), C, C++, JAVA, SQL, HTML/CSS, BASH, Javascript
ML/DL Libraries: PyTorch, Tensorflow, OpenCV, NLTK, Gensim, Huggingface-Transformers, OpenAI Gymnasium
GenAI Tools: LangChain, OpenAI, vLLM, SGLang, Unsloth, LLAMA-Factory
Retrieval/Database Tools: MySQL, Pinecone, MongoDB, FAISS, LlamaIndex, Recbole
Other Tools and Frameworks: Docker, Git, Linux/UNIX, Flask, Streamlit, Django, REST, Aether-Client