

ARNAV GOEL

+1-(412)-660-2046 | [✉ arnavgoe@cs.cmu.edu](mailto:arnavgoe@cs.cmu.edu) | [Google Scholar](https://scholar.google.com/citations?user=arnav10goel) | [🔗 arnav10goel.github.io](https://github.com/arnav10goel) | [in linkedin.com/in/arnavg10](https://www.linkedin.com/in/arnavg10) | [🐙 Github](https://github.com/arnav10goel)

EDUCATION

Carnegie Mellon University

Pittsburgh, PA

Master of Science in Machine Learning, Machine Learning Department (SCS) – **GPA: 4.0**

Aug 2025 - Dec 2026

- **Coursework:** Advanced Machine Learning, Deep Learning Systems, Multimodal Machine Learning, Diffusion and Flow Matching
- **Ongoing Research:** Investigating RL for efficient LLM and agentic reasoning with [Dr. Daniel Fried](#), [Dr. Sean Welleck](#) and Amazon AGI.

Indraprastha Institute of Information Technology (IIIT) Delhi

New Delhi, India

B.Tech in Computer Science and Artificial Intelligence – **CGPA: 9.50/10.00 (Class Rank: 2)**

Sept 2021 - June 2025

- **Coursework:** Data Structures, Algorithm Design, Database Management Systems, Object-Oriented Programming, Computer Arch., Operating Systems, Machine Learning, Natural Language Processing, Information Retrieval, Deep Learning, Reinforcement Learning

TECHNICAL SKILLS

Languages: Python (Numpy, Pandas, Matplotlib, Scikit-Learn, Scipy), C, C++, JAVA, SQL, HTML/CSS, BASH, Javascript

Tools and Frameworks: (1) **ML/DL Libraries:** PyTorch, NVIDIA NeMo, OpenCV, NLTK, Gensim; (2) **GenAI Tools:** LangChain, OpenAI, Huggingface-Transformers, vLLM, Unsloth, Apache Spark; (3) **Retrieval/Database Tools:** MySQL, Pinecone, MongoDB, FAISS, LlamaIndex; (4) **Other Tools:** Slurm, Docker, Git, Linux/UNIX, Flask, Streamlit, Django, REST, Aether-Client, Azure ML

PUBLICATIONS (ALL FIRST/CO-FIRST AUTHOR)

1. **HORIZON: A Benchmark for In-the-wild User Behaviour Modeling** [[NeurIPS 2025 & ACL 2026](#)]
2. **Attributing Culture-Conditioned LLM Generations to Pretraining Corpora** [[ICLR 2025](#)]
3. **Optimizing Vision Language Models for Scientific VQA through Caption-Aware Supervised Training** [[ACM-MM 2025](#)]
4. **Exploring Multilingual Speaker Emotion Recognition: Leveraging Co-Attention Cues in Multitask Learning** [[INTERSPEECH 2024](#)]
5. **Multilingual Prosody Transfer: Comparing Supervised & Transfer Learning** [[ICLR 2024](#)]
6. **LLMGuard: Guarding Against Unsafe LLM Behavior** [[AAAI 2024](#)]

WORK EXPERIENCE

Amazon Web Services (AWS)

May 2026 - August 2026

Applied Scientist Intern | AWS Agentic AI

Seattle, WA

- Improved **instruction following** in **multi-agent systems** by developing **context routing** strategies for efficient information selection and agent coordination, increasing task reliability across complex enterprise workflows.
- Enhanced agent reliability by integrating **verification** through **autoformalization** and **LLM-based judges**, and trained **domain-specific policy models** using **GRPO** to optimize agent behavior for analytics tasks.

Microsoft Research (MSR)

Jan 2025 - July 2025

Research Engineering Intern | Supervisors: [Dr. Amit Sharma](#) (Extreme Classification Group)

Bangalore, India

- Built **HORIZON**, a large-scale user modeling benchmark from long-range cross-domain histories, with optimized big-data handling (**Apache Parquet**, **SQLite**). Proposed a novel **FAISS**-based dual-encoder retrieval setup to evaluate **LLM-based recommenders**.
- 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 and training using **distributed multi-GPU** setups (**vLLM** & **Axolotl**). Accepted to **NeurIPS-2025 Workshop**.
- 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.

IBM Research

May 2023 - Oct 2023

Research Engineering Intern | Supervisors: [Dr. Sameep Mehta](#), [Dr. Nishtha Madaan](#)

New Delhi, India

- Developed **lightweight LLM guardrails** for **safer** enterprise applications which **blacklist user-suggested topics (87% accuracy)** via LDA & keyword extraction) & **detect toxicity (91–95% with a bi-encoder)**, and **hallucinations** (RAG-based response grounding).
- 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**.

RESEARCH EXPERIENCE

University of Southern California

May 2024 - Jan 2025

IUSSTF-Viterbi ML Research Intern | Supervisor: [Dr. Xiang Ren](#)

Los Angeles, CA

- Designed MEMOED, the first framework attributing LLM cultural knowledge to **memorization vs. generalization**, revealing influence of **pre-training data distributions** and enabling **corpus quality and bias assessment** for **LLM safety**. ([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**.

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](#).

PROJECTS

- **Learning to Act without Actions (ICLR 2024):** Built a system that learns **latent action policies directly from videos (LAPO)**, enabling pre-training of **generalist reinforcement learning agents** without action-labeled data. || [GitHub](#)
- **Amazon ML Challenge'24:** Designed a novel **few-shot learning** paradigm over **multimodal LLMs (VLMs)** and **ensembled open-source VLMs** for entity extraction from e-commerce images, securing **6th rank in India** with a **0.72 F1 score**. || [GitHub](#)
- **Multimodal Browser Agent:** Developed a JavaScript & CSS-based Chrome extension with real-time web scraping, OOPs principles, **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**. || [GitHub](#)

ACHIEVEMENTS

- **Google DeepMind Research Symposium 2025:** Selected among top 150 undergraduates in India to attend the symposium by GDM.
- **Amazon ML Challenge 2024:** Ranked **6th** out of **20,000+** teams nationwide, securing an **Applied Science Internship at Amazon**.
- **IUSSTF-Viterbi India Program 2024:** Selected among **15 nationwide scholars** for funded research internship at **USC-NLP group**.
- **MITACS Globalink Research Scholarship 2024:** Conducted fully-funded NLP internship at **MILA-Quebec AI Institute**.