

# Rahul Chand

650-509-9551 | [rahulsc@stanford.edu](mailto:rahulsc@stanford.edu) | [chandrahul0320@gmail.com](mailto:chandrahul0320@gmail.com)

Website: [rahulschand.github.io](https://rahulschand.github.io) | LinkedIn: [RahulSChand](https://www.linkedin.com/in/RahulSChand) | GitHub: [RahulSChand](https://github.com/RahulSChand)

## EDUCATION

<b>Stanford University</b> Masters in Computer Science   4.0 GPA	2024-2026
<b>Birla Institute of Technology, Pilani</b> Bachelor in Computer Science   9.7 GPA	2015-2019

## INDUSTRY AND RESEARCH EXPERIENCE

<b>NVIDIA   Intern @ GEAR Lab</b> • Working in the GEAR lab on foundational models for humanoid robots (Gr00t N1).	Sept 2025 -
<b>NVIDIA   Intern @ Nemotron Post-training</b> • Worked in the Nemotron LLM post-training team on exploring latent reasoning and async RL for nemo-rl.	June 2025 - Sept 2025
<b>Stanford AI Lab (ILIAD)</b> • Working as a grad research student	Nov 2024 -
<b>Microsoft Research   Pre-Doctoral Fellow</b> • Worked on Extreme Multi Label Learning (multi-label classification with million+ labels). I primarily worked on improving performance of extreme classifiers via distillation. <a href="#">ICLR'24</a> • Worked with Turing Team on factorization and sparsity methods for Transformer compression. <a href="#">EMNLP 2022</a>	July 2021 - Nov 2023
<b>Arcesium (De-Shaw)   SWE</b> • Developed microservices & frontend using Kotlin, React & SQL for handling large volumes (>100k) of trades daily.	Sept 2019 - March 2021
<b>Indian Institute of Science (IISc)   Intern</b> • Worked at Video Analytics Lab (VAL), IISc on using capsule networks for optical flow estimation. <a href="#">arxiv</a>	Jan 2019 - July 2019

## PUBLICATIONS

- Importance Weighted Retrieval for Few-shot Imitation Learning (**CorL 2025 Oral**) ([Link](#))  
*Amber Xie, Rahul Chand, Dorsa Sadigh, Joey Hejna*
- Enhancing Tail Performance in Extreme Classifiers by Label Variance Reduction (**ICLR 2024**) | ([Link](#))  
*Anirudh Buvanesh, Rahul Chand (*co-first author*), Yashoteja Prabhu, Manish Gupta*
- DSFormer: Effective Compression of Text-Transformers by Dense-Sparse Weight Factorization ([Link](#))  
*Rahul Chand, Yashoteja Prabhu, Pratyush Kumar.*
- CapsFlow: Optical Flow Estimation with Capsule Networks ([Link](#))  
*Rahul Chand, Rajat Arora, Ram Prabhakar, Venkatesh Babu.*

## OPEN SOURCE CONTRIBUTIONS

- Compute requirements for LLMs** ([Github](#)) | 1400+ ★
  - Tool to check GPU vRAM requirement and token/s for training & inference of any LLM. Supports frameworks like HuggingFace, vLLM, llama.cpp, and quantization (bitsandbytes, GGML & QLoRA).
  - The tool has been used over **150k+ times** by **25k+** users.
- llama2.c for dummies** ([Github](#)) | 200+ ★
  - Walkthrough tutorial of [llama2.c](#) written as a starter reference for LLM inference.
- Nemo-RL** ([Github](#)) | 1000+ ★ (contributor)
  - Contributed to the Nemo-RL post-training library during my internship in the Nemotron LLM team. I worked on async rl
- Nemo-Gym** ([Github](#)) | 70+ ★ (contributor)
  - Contributed to the Nemo-Gym RL environment library

---

## TEACHING EXPERIENCE

---

### TA (Stanford, MS)

- Winter 2025: CS153

### TA (BITS Pilani, BS)

- Fall 2018: Data Mining, Principles of Programming Languages, C Programming
- Spring 2018: Data Structures & Algorithms, Database Systems
- Fall 2017: Logic in Computer Science