

# RAGHAV KAPOOR

ML/NLP Graduate Student at Carnegie Mellon University, School of Computer Science

☎ +1-412-692-0131

✉ raghavka@cs.cmu.edu

📧 raghav-kapoor

🏠 raghav1606.github.io

in raghavkapoor16

🔗 Raghav1606

## EDUCATION

- Carnegie Mellon University, School of Computer Science** Pittsburgh, PA  
*Masters, Intelligent Information Systems (ML/NLP Specialization), GPA: 4.17/4.0* [Aug 2022 - (Exp) Dec 2023]  
**Advisors:** Dr. Graham Neubig, Dr. Yonatan Bisk | **Teaching Assistant:** Search Engines (Dr. Jamie Callan)  
**Courses:** Advanced NLP, Multimodal Learning, Search Engines, Introduction to ML, Advanced DL, Cloud Computing
- Netaji Subhas Institute of Technology, Delhi University** New Delhi, India  
*Bachelor of Engineering - Computer Engineering; GPA: 9.11/10.0 (Department Rank: 5)* [Aug 2015 - May 2019]  
**Courses:** Data Structures, Analysis Of Algorithms, Networking, Operating Systems, Databases, Artificial Intelligence

## PUBLICATIONS

- EXCALIBUR: Evaluating and Encouraging Embodied Exploration**  
H Zhu, R Kapoor, SY Min, W Han, J Li, K Geng, G Neubig, Y Bisk, A Kembhavi, L Weihs. **CVPR 2023** [Accepted]
- In-Context Learning with GPT-4 for Medical Summarization**  
Y Mathur\*, R Kapoor\*, S Rangreji\*, M Palavalli\*, A Bertsch, M Gormley **ACL ClinicalNLP 2023** [paper]
- Hate Me Not: Detecting Hate Inducing Memes in Code Switched Languages**  
K Rajput\*, R Kapoor\*, K Rai\*, P Kaur. **AMCIS 2022** [paper]
- Transfer Learning for Detecting Hateful Sentiments in Code Switched Language**  
K Rajput, R Kapoor, P Mathur, P Kumaraguru, RR Shah. Deep Learning Based Approaches for Sentiment Analysis. Springer 2020 [paper]
- Mind your language: Abuse and offense detection for code-switched languages**  
R Kapoor, Y Kumar, K Rajput, Rajiv R Shah, P Kumaraguru, R Zimmermann, **AAAI 2019** [Best Student Poster] [paper]

(\* Denotes equal contribution)

## RESEARCH EXPERIENCE

- NeuLab and CLAW Lab - Carnegie Mellon University** Pittsburgh, PA  
*Student Researcher - Multimodal ML, Embodied AI, Reinforcement learning, VR/AR* [Aug 2022 - Present]
  - **Advisors:** Dr. Graham Neubig, Dr. Yonatan Bisk, **PRIOR @ Allen Institute of Artificial Intelligence (AI2)**
  - Implemented human-level features of the AI2THOR in randomized environments in Virtual Reality (VR) using Oculus.
  - Includes replicating continuous movement to discretized steps in VR to human-arm movement in real-world, mapping force to Oculus grip, Opening and Closing objects in VR, and rendering procedurally generated randomized scenes in VR.
  - Built and trained RL agents for all phases using Proximity Policy Optimization Algorithm (PPO) and Semantic Maps.
  - Developed end-to-end pipeline for QA generation using scene graphs and run baselines for exploration and QA. [video]
- MIDAS Lab, IIIT Delhi** New Delhi, India  
*Research Intern - Natural Language Processing, AI for Social Good* [Aug 2018 - Jun 2019]
  - **Advisors:** Dr. Rajiv Ratn Shah, Dr. Ponnurangam Kumaraguru, Dr. Roger Zimmermann
  - Invented an LSTM cum transfer learning model for Hinglish offensive text classification (Best poster award in AAAI)
  - Researched and trained own word embeddings for code-mixed languages (Achieved state of the art accuracy: 87%)
  - Led the dataset curation, annotation and cleaning initiative in the lab for 5000+ hinglish tweets for model pipeline. [video]

## HONORS AND AWARDS

- Winner**, Smart India Hackathon' 19, Govt. of India - For Project PropSafe (Among 6000+ teams) [2019]
- Finalist (Top 10)**, Goldman Sachs Machine Learning Hackathon'19 (Among 100+ Teams) [2019]
- Best Student Poster Award and Travel Grant**, For Paper: Mind Your Language at AAAI, Hawaii [2019]
- Honourable Mention**, ACM ICPC-18 Gwalior-Pune Regionals [2018]
- Grand Finalist**, Smart India Hackathon'18, Govt. of India - For Project UPCHAAR (Among 5600+ teams) [2018]
- College Merit Scholarship**, awarded twice for Academic Excellence (NSIT). Ranked in top 5% in department [2017-18]
- Qualified for Indian National Astronomy Olympiad**, HBCSE (Among top 300 or 1%ile nationwide) [2014]

## INDUSTRY EXPERIENCE

---

- **Adobe Inc.** San Jose, CA  
*Intern, Software Engineer | Adobe Document Sign* [May 2023 - Present]
  - Building the bulk contract features including bulk signing, and enabling sign field detection for 100k+ contracts at once.
  - Developed a model for auto signature generation and authenticate sign and identity using certificates and user data.
- **Goldman Sachs** Bengaluru, India
  - *Associate, Software Engineer | GS Electronic Trading Platforms* [Jan 2022 - Jul 2022]
    - Built and managed nanosecond latency trading systems for large volume stocks and options flow worth billion dollars.
    - Developed a tool using Java to effectively distribute short sell locates to clients, increasing daily trade volume by 35%.
    - Drove initiatives with external vendors in US exchange connectivity space to enhance trade liquidity with darkpools.
  - *Analyst, Software Engineer | GS Electronic Trading Platforms* [Jun 2019 - Dec 2021]
    - Managed load capacity planning and effectively redesigned the order management system to cut latency by 40%.
    - Designed an ML/NLU based chatbot for traders to answer queries from traders, reducing human dependence by 60%.
  - *Summer Analyst Intern, Software Engineer | GS Electronic Trading Platforms* [May 2018 - Jul 2018]
    - Developed a web-based monitoring solution using Python, Flask, Shell that curated all host-related information from single API and performed health check for hosts to enhance resiliency.

## PROJECTS

---

- *LLMs for Action Model Learning*
  - Building an LLM for action model learning, i.e, automated human-computer interaction, for technically challenged people.
  - Perform basic tasks like ordering food, booking flight tickets, and replying to mails through singular text/voice commands.
  - Curating data and training models for automated understanding of UX components. Fusing knowledge from existing LLMs to create a robust solution[\[link\]](#)
- *Search Engine*
  - Built a search engine for over 4 million documents in Java that supports boolean retrieval, BM25, Indri retrieval algorithms.
  - Implemented query expansion and Pseudo Relevance Feedback. Implemented diversification using PM2 and xQuAD.
  - Added ML-based reranking capabilities using custom features, learning to rank (LTR) like ListNet, Coordinate Ascent.[\[link\]](#)
- *MaRVLous*
  - Performed extensive analysis of existing methods on the MaRVL dataset on visual-linguistic Inference for multilingual domains.
  - Built a pipeline for Domain adaptive CCLM model involving subsampling relevant examples from WiT using Faiss library and training further using multi-task learning similar to GiVL[\[link\]](#)
- *Dvesh Prahari: Attackers in Disguise (B.E. Project Thesis)*
  - Deals with detecting hate speech and sarcasm in political memes and twitter pages in context of code-switched languages
  - Curated and annotated an IPM (Indian Political Memes) dataset and used code-switched embeddings over CNN-LSTM model.
  - Applied the model to several twitter pages to filter out the twitter pages that provoke vindictive sentiments in elections[\[link\]](#)
- *PropSafe*
  - A web application designed to automate the land/real estate transaction system used by land registry offices in India.
  - The app detects property frauds using AI based checks and smart contracts/blockchain for maintaining tamper-free records
  - Used random forest and SVM for tabular data pertaining to a transactor's background, and Truffle for record retentions[\[link\]](#)

## SKILLS AND CERTIFICATIONS

---

- **Languages:** Python, C/C++, Shell Scripting, Java, Javascript, C#, SQL
- **Technical Skills:** Research, Machine Learning, Natural Language Processing, Multimodal ML, Web Development
- **IT Constructs:** Cloud Computing, Distributed Systems, ETL, Linux, Data Structures, Algorithms, DBMS, System Design
- **Libraries/Frameworks:** Pytorch, OpenAI Gym, Unity3D, Ai2Thor, Lucene, Scikit, Flask, Node.js, React, Groovy
- **Certifications:** Deep Learning Specialization by AndrewNg; Machine Learning - Coding Blocks; Udemy Java Certification

## VOLUNTEERING EXPERIENCE

---

- **Conference Reviewer**, Participated in blind review process for ACL Clinical NLP Workshop 2023
- **Vice President and Club Mentor**, GS Toastmasters Club, Handled member concerns, maintained influx of new members into the club. Mentored 40+ members. Won Triple Crown, Pioneer Gold Award for public speaking.
- **Senior Member and Educator**, Algorithm Society of College, Taught algorithmic concepts to 60+ students weekly.
- **Volunteer**, Community TeamWorks at Goldman Sachs, to create educational material for intellectually challenged kids.