

# Sameer Jain

Email: [jainsameer0612@gmail.com](mailto:jainsameer0612@gmail.com)

LinkedIn: [sameerjain06](#)

Phone: +1 412-983-8708

## EDUCATION

---

### Carnegie Mellon University, School of Computer Science

Pittsburgh, PA

Master of Science in Intelligent Information Systems (Language Technologies Institute)

December 2022

- Coursework: Introduction to ML, ML with Large Datasets, Human Language for AI, Advanced NLP, Multilingual NLP, Computational Ethics for NLP, Probability and Mathematical Statistics

### Birla Institute of Technology and Science

Pilani, India

Bachelor of Engineering (Hons.) Computer Science (Distinction)

June 2019

- Thesis: *Copy Mechanisms for Upstream Text Processing* [\[pdf\]](#)

## EXPERIENCE

---

### Supernormal

Mountain View, CA

Machine Learning Engineer | AI Core

April 2023 - Present

- Supernormal captures meetings and generates notes, summaries, and action items using large language models (**LLMs**)
- Lead and own **note quality evaluation** and prompt engineering for note quality enhancement

### Carnegie Mellon University

Pittsburgh, PA

Research Assistant (*Mentors: Fei Fang and Lei Li*)

March - June 2023

- Partnered with the World Wide Fund for Nature to identify **multilingual** environmental conservation news
- Designed and deployed a system utilizing large language models (**LLMs**) to identify conservation news in low-resource languages such as Nepali
- Achieved state-of-the-art performance using **< 10%** of the data required by existing methods, significantly reducing manual annotation efforts.

### Meta

Menlo Park, CA

Software Engineering Intern | Ad Enhancement and Delivery

June - August 2022

- Collaborated with the Creative Engineering team on augmented reality (AR) enhanced ad transformations
- Built **pipelines** to identify best **AR effects for ad enhancement** and built guardrails against low-quality transformations

### Samsung Research Institute

Bangalore, India

Software Engineer | Voice Intelligence R&D

July 2019 - December 2020

- Developed **on-device** language capabilities to reduce server costs and improve offline performance of Bixby
- Designed memory-efficient data structures and caching mechanisms to improve latency of the on-device NLU engine

### University of Zurich

Zurich, Switzerland

Research Intern (*Mentor: Tanja Samardžić*)

January - May 2019

- Improved **low-resource** performance of LSTM encoder-decoders on morphological tasks by introducing a **copy mechanism**
- Applied the architecture for normalization of the Swiss-German language and for morphological segmentation of the English, German, and Indonesian languages

### Aix-Marseille University and CNRS

Aix-en-Provence, France

Research Intern (*Mentor: Philippe Blache*)

May - July 2018

- Designed a model to **automatically estimate presence** during humans interactions with embodied conversational agents
- Built **multimodal** classifiers to estimate this from humans' verbal and non-verbal cues during virtual reality interactions, significantly **reducing human effort** as presence estimation is typically performed using subjective questionnaires

## PUBLICATIONS

---

1. “Where It Really Matters: Few-Shot Environmental Conservation Media Monitoring for Low-Resource Languages”  
**Sameer Jain**, Sedrick Keh, Shova Chettri, Karun Dewan, Pablo Izquierdo, Johanna Prussman, Pooja Shreshtha, César Suárez, Ryan Shi, Lei Li, and Fei Fang  
*In submission to AAAI '24* **Preprint**
2. “Multi-Dimensional Evaluation of Text Summarization with In-Context Learning” [\[pdf\]](#)  
**Sameer Jain**, Vaishakh Keshava, Swarnashree Mysore Sathyendra, Patrick Fernandes, Pengfei Liu, Graham Neubig, and Chunting Zhou  
*Findings of the Annual Meeting of the Association for Computational Linguistics* **ACL Findings '23**
3. “Four Years of FAccT: A Reflexive, Mixed-Methods Analysis of Research Contributions, Shortcomings, and Future Prospects” [\[pdf\]](#)  
Benjamin Laufer, **Sameer Jain\***, A. Feder Cooper\*, Jon Kleinberg, and Hoda Heidari  
*Proceedings of the ACM Conference on Fairness, Accountability, and Transparency* **FAccT '22**
4. “Toward an Automatic Prediction of the Sense of Presence in Virtual Reality Environment” [\[pdf\]](#)  
Magalie Ochs, **Sameer Jain**, and Philippe Blache  
*International Conference on Human-Agent Interaction* **HAI '18**

## PROJECTS

---

### Evaluation of Machine-generated Text using In-context Learning

CMU

*Mentors: Chunting Zhou and Graham Neubig*

August - December 2022

- Designed a framework for using **large language models (LLMs)** to **evaluate the quality** of artificially generated summaries
- Achieved **state-of-the-art** performance on quality evaluation while using **< 0.1%** of the training data used by existing methods

### Mixed-method Reflexive Analysis of FAccT

CMU

*Mentors: Hoda Heidari and Jon Kleinberg*

January - December 2021

- Identified areas of focus in the **fair machine learning** domain through community detection in fair-ML citation networks
- Uncovered the moral underpinnings of Fair-ML literature against the framework given by the Moral Foundations Theory

## SKILLS

---

- **Languages:** Python, C/C++, Java, SQL
- **Frameworks:** PyTorch, TensorFlow, Spark

## SCHOLARSHIPS AND AWARDS

---

- **National Talent Search Examination (NTSE) Scholarship**, awarded to **1000 students among 1 million** candidates by the National Council of Educational Research and Training (NCERT), India
- **Kishore Vaigyanik Protsahan Yojana (KVPY) Fellowship**, offered to **800 candidates out of 50,000 applicants** by the Department of Science and Technology, Government of India, in collaboration with Indian Institute of Science, Bangalore
- All India Rank **381 among 1.3 million candidates** in **IIT-JEE Mains 2015**
- All India Rank **1606 among 117,000 candidates** in **IIT-JEE Advanced 2015**
- **BITS Pilani Merit Scholarship**, awarded to the **top 3%** of the class in the 2017-18 academic year