Raj Sanjay Shah

✉ rajsanjayshah@gmail.com 📞 6282809784 📇 TSRB 228, 85 5th St NW, Atlanta, GA 30332

in www.linkedin.com/in/raj-sanjay-shah Raj-Sanjay-Shah 🌐 `https://raj-sanjay-shah.github.io/`

About me

Ph.D. student at Georgia Institute of Technology. As AI becomes pervasive, I want to pursue a Ph.D. program with a focus on understanding the human behavioral alignment of language models and the impact on social decision-making. During my Master's program at Georgia Tech, I have been fortunate to work on a variety of projects that use language processing for the betterment of online mental health platforms. This has equipped me with a strong foundation in computational social science methods and a well-rounded experience working with popular language processing algorithms.

Education

Aug 2023 –	**Georgia Institute of Technology** *Ph.D in **Computer Science*** Advisors: Professor Sashank Varma (Georgia Tech), Professor Diyi Yang (Stanford)
Aug 2021 – May 2023	**Georgia Institute of Technology** *MS in **Computer Science** (GPA: 4.0/ 4.0, Specialization: Machine Learning)*
Aug 2017 – Jul 2021	**Birla Institute of Technology and Science, Pilani, India** *Bachelor of Engineering in Computer Science (CGPA: 8.54 / 10)* *Minor in Data Science (Minor GPA: 9.8 / 10)*

Employment

May 2023 – Aug 2023	**IBM Research - Human-Centered AI Team** *Sr. Research Software Engineer - AI* Explored different intervention strategies for high-quality human-AI collaborative data creation. Built a full-stack application for user studies. User studies investigate the idea of appropriate user reliance on AI agents using subtle real-time nudges.
May 2022 – Aug 2022	**Epic Systems -** *Software Developer - ML Intern* Trained Large Language Models to predict workflow actions of Health Care Providers with the goal of bringing relevant patient information to the Provider Dashboard on-the-fly.
Jan 2021 – Jun 2021	**JPMorgan Chase -** *Software Development Intern* Worked in the Merchant Services team on developing automated testing scripts for the onboarding platform using the Postman tool.
May 2020 – Jun 2020	**Samsung R&D Institute, India -** *Research Intern* Developed a bidirectional Long-short term memory based variable autoencoder to generate melody and chords for piano music as a part of the User-Experience: Ringtone project.
May 2019 - Jul 2019	**Mumbai Railway Vikas Corporation, India** *Software Intern* Created an android and web based Project Management System for efficient intra-company communication.

Awards

Georgia Tech President's Fellowship
This award is given to students who have excelled in scholarship and are pursuing doctoral degrees. Selection for this competitive award is based on multiple areas including but not limited to academic history, standardized test scores, and letters of recommendation. This recognition comes with a stipend of $5,500, per academic year, renewable up to 3 years.

Georgia Tech Students Honors Program - The Marshall D. Williamson Fellowship Award 2023:
Awarded to a well-rounded, second-year Master's student who best embodies Marshall's values of academic excellence and leadership.

Georgia Tech Students Honors Program - Donald V. Jackson Fellowship Award 2022:
Awarded to a well-rounded, first-year Master's student who best embodies the values of academic excellence and leadership.

Birla Institute of Technology and Science best paper award:
Awarded to computer science undergraduate for rigor in research and academics.

Lalita and Ravindra Nath (LRN) Foundation for student travel grant:
Awarded for travel to EMNLP 2022. The award is given to selected student candidates who have an accepted paper as a first author at a top tier International conference outside India and also have a paper or poster in COMSNETS main conference or one of the associated workshops (in any of the previous COMSNETS).

Publications

2023

1. Hsu, S.-L. *et al. Helping the Helper: Supporting Peer Counselors via AI-Empowered Practice and Feedback* 2023. arXiv: 2305.08982 [cs.HC].

2. **Shah, R. S.**, Marupudi, V., Koenen, R., Bhardwaj, K. & Varma, S. *Numeric Magnitude Comparison Effects in Large Language Models* in *Findings of the Association for Computational Linguistics: ACL 2023* (Association for Computational Linguistics, Toronto, Canada, July 2023), 6147–6161.

2022

4. Pathak, A., **Shah, R. S.**, Kumar, V. & Jakhotiya, Y. JARVix at SemEval-2022 Task 2: It Takes One to Know One? Idiomaticity Detection using Zero and One Shot Learning. *CoRR* **abs/2202.02394.** arXiv: 2202.02394 (2022).

5. **Shah, R. S** *et al. Modeling Motivational Interviewing Strategies On An Online Peer-to-Peer Counseling Platform* in *Proceedings of the ACM on Human-Computer Interaction 6* (2022).

6. **Shah, R. S** *et al. When FLUE Meets FLANG: Benchmarks and Large Pretrained Language Model for Financial Domain* in *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (Association for Computational Linguistics, 2022).

2021

7. **Shah, R. S.**, Bhatia, A., Gandhi, A. & Mathur, S. *Bitcoin Data Analytics: Scalable techniques for transaction clustering and embedding generation* in *2021 International Conference on COMmunication Systems NETworkS (COMSNETS)* (2021), 1–6.

2020

8. Kristiansen, L.-M., Agarwal, V., Franke, K. & **Shah, R. S.** *CTI-Twitter: Gathering Cyber Threat Intelligence from Twitter using Integrated Supervised and Unsupervised Learning* in *2020 IEEE International Conference on Big Data (Big Data)* (2020), 2299–2308.

9. **Shah, R. S.** & Bhatia, A. *Bitcoin Data Analytics: Exploring Research Avenues and Implementing a Hadoop-Based Analytics Framework* in *Web, Artificial Intelligence and Network Applications* (eds Barolli, L., Amato, F., Moscato, F., Enokido, T. & Takizawa, M.) (Springer International Publishing, Cham, 2020), 178–189.

Leadership and Teaching Experience

Jan 2023 - May 2023	**CS 3790: Intro to Cognitive Science, Georgia Tech** *Graduate Teaching Assistant (Head TA) - Professor Sashank Varma* • Created homework and assignments for the course CS 3790.
Aug 2022 - Dec 2022	**CS 4650: Natural Language Processing, Georgia Tech** *Graduate Teaching Assistant - Professor Alan Ritter* • Making coding exercises for students to have hands on NLP experience.
Aug 2021 - Dec 2021	**CS 3790: Intro to Cognitive Science, Georgia Tech** *Graduate Teaching Assistant - Professor Sashank Varma* • Created homework and assignments for the course CS 3790.
Aug 2020 - Dec 2020	**CS F469: Information Retrieval, Birla Institute of Technology and Science, Pilani** *Undergraduate Teaching Assistant - Professor Vinti Agarwal* • Created a term paper-based assignment to promote student exposure to research.
Aug 2019 - Dec 2019	**Academic Counselling Cell, BITS Pilani** *Undergraduate Teaching Assistant - Professor Sangeeta Sharma* • Planned mentorship programs, development seminars, and community bonding activities.

Ongoing Research Projects

Aug 2022	**Professor Sashank Varma -** *Part of the Cognitive Architecture Lab at Georgia Institute of technology* - Performing behavioral bench-marking of NLP models by observing Numeric Magnitude Comparison Effects in Large Language Models - Mitigating catastrophic forgetting in Artificial Neural Networks by simulating a distributed power-law training environment - Combining Vision and Language Models Better Accounts for the Typicality Effect in Human Conceptual Understanding - Pre-training LLMs using human-like development data corpus
Aug 2021	**Professor Diyi Yang -** *Part of the SALT LAB at Stanford/ Georgia Institute of technology* - CARE: Empowering Peer Counselors via Automatic Suggestion Generation Submitted to CHI 2024 - AI for Mental Health: What Makes Digital Support Effective? How Therapeutic Skills Affect Clinical Well-Being Submitted to CSCW 2024
May 2021	**Professor Jordan Boyd-Graber -** *Intern at University of Maryland, College Park* QAML (Question-writing Aided by Machine Learning) is an interface for adversarial question-writing enhanced by the QANTA QA dataset.

Service

Reviewer for the following conferences

Association of Computational Linguistics 2024, Computer-Supported Cooperative Work And Social Computing 2023, ACM CHI Conference on Human Factors in Computing Systems 2023, Empirical Methods in Natural Language Processing 2022 (Computational Social Science and Cultural Analytics Track), IEEE BigData 2021.

Student Volunteer for the following conferences

Computer-Supported Cooperative Work And Social Computing 2022, 2021; Annual Meeting of the Association for Computational Linguistics 2021.

Mentorship

- Mentoring experiences with student collaborators:

- Atith Gandhi - MS CS @ GT (January 2023 -)
- Khushi Bhardwaj - Rising Junior @ GT (August 2022 -)
- Harsh Lalai - Rising Junior @ Bits Pilani, India (August 2022 -)
- Siddhartha Vemuri - MS CS @ GT (January 2023 - August 2023)
- Jason Liu - Undergraduate student @ UPenn (August 2021 - August 2022)
- Damian Rene - Undergraduate student @ UPenn (August 2021 - August 2022)

- Student mentor and judge at Undergraduate Research Opportunities Program(UROP) @GT

Spoken Languages

English (Native Proficiency) , Hindi (Native Proficiency), Gujarati (Native Proficiency), Marathi (Expert Proficiency), Sanskrit (Intermediate Proficiency). I have also completed a MOOC titled Miracles of Human Languages: An Introduction to Linguistics (link to the certificate).