

# Pratik JOSHI

[pratikmjoshi.github.io](https://pratikmjoshi.github.io) [Google Scholar](https://scholar.google.com/citations?user=pratikmjoshi) [in linkedin.com/in/pratik-joshi-645305150](https://www.linkedin.com/in/pratik-joshi-645305150) [@ pratikmjoshi123@gmail.com](mailto:pratikmjoshi123@gmail.com)  
[+1 412-608-7330](tel:+14126087330) [github.com/pratikmjoshi](https://github.com/pratikmjoshi)

## EDUCATION

- 
- May 2023 | **CARNEGIE MELLON UNIVERSITY - SCHOOL OF COMPUTER SCIENCE, Pittsburgh, PA, USA**  
Master of Science in Intelligent Information Systems, Language Technologies Institute
- Jun. 2019 | **BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE PILANI, Goa, India**  
Bachelor of Engineering (Honors) in Computer Science | CGPA : 8.44/10.00  
*Relevant courses : Machine Learning, Data Mining, Neural Networks, Artificial Intelligence, Data Structures and Algorithms, Object-Oriented Programming, Computer Vision, Operations Research*

## SKILLS

---

**Languages** Python, C/C++, Shell Scripting, MySQL, Prolog (intermediate)  
**Toolkits** PyTorch, Keras, Tensorflow, Matplotlib, Android Studio (intermediate)

## EXPERIENCE

- 
- Sep. 2020 | **GOOGLE RESEARCH, Bangalore, India**  
Aug. 2021 | *Research Engineer (via Optimum Infosystem) | Manager : Dr. Partha Talukdar*
  - > Developed entity-aware translation architecture for multilingual subtitle generation of technical videos; received 9/10 score from manual raters and translated 100 hours of video content for an educational organization (NPTEL).
  - > Trained neural models for temporal parsing of multilingual queries and created benchmark for model evaluation.
- Aug. 2019 | **MICROSOFT RESEARCH, Bangalore, India**  
Aug. 2020 | *Research Intern | Managers : Dr. Kalika Bali, Dr. Monojit Choudhury*
  - > Probed natural language inference capabilities of pre-trained language models (BERT, RoBERTa). [CoNLL'20]
  - > Analyzed resource disparity for 2500 languages, and published strongly cited paper (80 citations). [ACL'20]
  - > Developed an annotation tool for code-mixed speech, which yielded 10% reduced error rate and increased efficiency over baseline tools. [EMNLP'19]
- Jul. 2018 | **MICROSOFT RESEARCH, Bangalore, India**  
Dec. 2018 | *Research Intern | Managers : Dr. Navin Goyal, Dr. Monojit Choudhury*
  - > Implemented neural semantic parsing techniques for converting natural language queries to code (Regex, SQL), which performed competitively with state-of-the-art techniques.
- May 2017 | **WHITE DATA SYSTEMS INDIA (I-LOADS), Chennai, India**  
Jul. 2017 | *Software Engineering Intern | Manager : Gokulan Jayaram*
  - > Created voice interface for flagship logistics app, and built command recognition for English and Hindi.

## SELECTED PUBLICATIONS

- 
- > Pratik Joshi\*, Sebastin Santy\*, Amar Budhiraja\*, Kalika Bali, Monojit Choudhury, "The State and Fate of Linguistic Diversity and Inclusion in the NLP World", **ACL'20** [pdf], Coverage : [NLP with Friends](#), [Ruder's blog](#), [Quartz](#), [LacunaFund](#), [SIGTYP Newsletter](#)
  - > Pratik Joshi\*, Somak Aditya\*, Aalok Sathe\*, Monojit Choudhury, "TaxiNLI : Taking a Ride up the NLU Hill", **CoNLL'20** [pdf]
  - > Sanket Shah, Pratik Joshi, Sebastin Santy, Sunayana Sitaram, "CoSSAT : Code-Switched Speech Annotation Tool", **EMNLP'19** [pdf]

## SELECTED PROJECTS

- 
- SEMANTIC QUESTION-ANSWERING SYSTEM** BITS PILANI | JUN. 2019
  - > Created Prolog implementation of the paper "Operation of a Semantic Question-Answering System" (Raphael 1963).
  - > Implemented numerical reasoning, positional inference, commonsense reasoning using relational entity graphs. [code]
- MULTI-DOCUMENT ABSTRACTIVE SUMMARIZATION** BITS PILANI | JUN. 2018
  - > Investigated and implemented different neural architectures for the summarization of the DUC corpus. Studied various encoder-decoder models, attention models and pointer networks. [code]