SAYALI GHODEKAR

smg2280@columbia.edu ◇ LinkedIn ◇ GitHub ◇ Google Scholar

EDUCATION

Columbia University *(MS in Computer Science)* *Expected Dec 2022*
Relevant Coursework: NLP, Speech Recognition, Adv Spoken language processing, Deep learning for Computer Vision, AI, ML and Climate.

Pune Institute of Computer Technology *(BE in Computer Engineering)* *June 2019*
Relevant Coursework: Machine Learning, Data Analytics, AI and Robotics, Soft Computing and Optimization, Engineering Mathematics, Cloud Computing, High Performance Computing. **GPA: 3.97/4**

WORK EXPERIENCE

RingCentral Inc, *Machine learning engineer (Conversational AI Team)* *Dec 2020 - July 2021*

- Created innovative Machine learning strategies for NLP problems, leading to integration of AI-based services in RingCentral's video communication platform for 300K businesses worldwide.
- Built, deployed and scaled abstractive summarization service to extract key insights from conversations, decreasing system latency. Implemented data collection, models and client libraries, dockerization, and Kubernetes deployments.
- Improved qualitative performance of post-meeting tools through text-generation services including Coreference-Resolver, question-generator, text-paraphraser, and multilingual summarizer.

Consumer Reports, *Data Science Intern* *May 2022 - Aug 2022*

- Created a pre-processing pipeline for cleaning noisy data of 50K brands scraped from CR's website using NLP and statistical techniques.
- Collaborated with the Product Safety team to design early warning system that integrates CR's 5 internal data channels using Twitter API and BERT Transformers.

DeepAffects (Conversation-Intelligence Startup), *Machine learning engineer* *Jan 2020 - Dec 2020*

- Developed and scaled DeepAffect's abstractive summarization API, leading to a 10% increase in model performance. Collaborated with quality team for continuous testing, processing over 1M+ minutes of audio into summaries.
- Enhanced DeepAffect's conversation metrics stack through dialogue act tagging, intent classification, question and answering systems, reducing time to market these APIs by 75%.
- Proposed and implemented a topic segmentation algorithm for multi-turn dialogue data, demonstrating an improvement in model performance across all NLP pipelines.

RESEARCH

Columbia NLP Lab *Aug 2022-*
Advised by Prof. Zhou Yu (in collaboration with Prof. Shinji Watanabe, CMU LTI)

- Working on an accented speech-recognizer to enhance Edubot, a language learning chatbot to help second-language learners.

Center for Indian Language Technologies, IIT Bombay *June 2019 - Dec 2019*
PI: Prof. Pushpak Bhattacharyya

- Designed experiments for Cognate Detection with Siamese neural networks, evaluating word embedding models and beating baseline by 71%.
- Collaborated with Cognitive NLP group to derive gaze features for cognate identification. Reduced annotation cost by leveraging predictive models, demonstrating a 12% improvement over baseline.
- Published and presented cognate study findings in ACM CoDS-COMAD 2021 [link] and EACL 2021 [link]. Awarded **best paper honorable mention at EACL 2021**.

TEACHING

Natural language processing, Fall 2021 *New York, NY*
Teaching Assistant for Graduate-level course at Columbia CS (Responsibilities: Weekly recitations, Office hours and Grading)

ACADEMIC PROJECTS

Automatic Speech Recognition System for Code-Switched speech *Sep 2021 - Dec 2021*

- Built an ASR system for recognizing Code-switched text for the low-resource Hindi-English language pair using the Kaldi toolkit. Created my self-recorded dataset and achived a WER of 36 on the test set.

TECHNICAL SKILLS

Python, pandas, numpy, C++, Torch, Tensorflow, Kaldi, scikit-learn, OpenCV, Docker, Kubernetes, AWS, Javascript, PHP, MySQL, MongoDB, PowerShell, Basics of Google Cloud Services.