

# SAYALI GHODEKAR

180 Claremont Avenue, New York, NY 10027

+1(929)677-0630 ◊ smg2280@columbia.edu ◊ GitHub ◊ LinkedIn

## EDUCATION

---

### Columbia University

#### *Master of Science, Computer Science*

Relevant Coursework: Fundamentals of Speech Recognition, Analysis of algorithms, Deep learning for Computer Vision, Databases.

*New York, NY*

*Expected Dec 2022*

### Savitribai Phule Pune University

#### *Bachelor of Computer Engineering*, GPA: **9.09/10**

Relevant Coursework: Machine Learning, Data Analytics, AI and Robotics, Soft Computing and Optimization, Engineering Mathematics, Cloud Computing, High Performance Computing.

*Pune, India*

*June 2019*

## WORK EXPERIENCE

---

### RingCentral Inc

#### *Data Scientist (Conversation AI Team)*

*Mumbai, India*

*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.

### DeepAffects - Conversation-Intelligence Startup

#### *Data Scientist*

*Mumbai, India*

*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.

### Center for Indian Language Technologies, IIT Bombay

#### *Research Intern under the guidance of Prof. Pushpak Bhattacharyya*

*Mumbai, India*

*June 2019 - Dec 2019*

- 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**.
- Conceptualized and designed novel Textual History Analysis Tool to capture historical evolution of texts using computational phylogenetics. Managed tool deployment 2 weeks prior to deadline, resolved technical conflict during presentation.

## TEACHING

---

### Natural language processing

#### *Teaching Assistant for Graduate-level course at Columbia Computer Science*

*New York, NY*

*Fall 2021*

## ACADEMIC PROJECTS

---

### Optical Character Recognition for Marathi Text with Braille Conversion Unit

#### *Final year project*

*Mumbai, India*

*June 2018 - Apr 2019*

- Led a team of four to build a novel, automated system to obtain Marathi text from images and further converts text in the Braille format using CNN-BiLSTM based model and achieved a 0.82 F1-score.
- Constructed a dataset of 12k Marathi word images with annotated labels for image processing and character recognition.

## TECHNICAL SKILLS

---

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