

Prachi Rahurkar

rahurkap@gmail.com
Web: prachirahurkar.github.io
LinkedIn: prachi-rahurkar
Mobile: (+1) 669-204-1088

EDUCATION

Oregon State University

Master of Science - Computer Science, AI Track, GPA: 3.71/4

Corvallis, OR

Sep 2018–Mar 2021

Thesis: Adversarial Attacks in Natural Language Question-Answering

Coursework: Natural Language Processing, Deep Learning, Machine Learning, Computer Vision- I and II, Networks in Computational Biology, Intelligent Agents and Decision-Making, Software Engineering

Thadomal Shahani Engineering College–University of Mumbai

Bachelor of Engineering - Computer Science, GPA: 8.26/10

Mumbai, India

Aug 2014–May 2018

Coursework: Analysis of Algorithms, Data Structures, Artificial Intelligence, Data Mining, Object-Oriented Programming, Computer Graphics, Distributed Databases, Human Machine Interaction, Cloud Computing

WORK EXPERIENCE

• Machine Learning Engineer

Amazon Music

Aug 2021–present

San Francisco, CA

- Designed and built customer embeddings training and validation jobs using track playback history data of Amazon Music Unlimited customers to generate personalized track recommendations (with a runtime of only 45-47 mins) along with its AWS clusters and CDK.
- Extended the above pipeline of same-entity (track-to-track similarity) use case to cross-entity (track-to-artist similarity) use cases, which was adopted by 4 client services enhancing customer listening experiences.
- Proposed and implemented sequencing logic to play first 5 tracks in playback personalized to the customer's taste when a Free-tier Music customer requests for a track; as Free customers cannot play songs on-demand. This improved customer engagement and retention (on Voice) on Amazon Music Free by 6%.
- Proposed and implemented the model to showcase personalized playlist MyDiscoveryMix which contains 15 never-heard-before tracks based on customer's listening history, increasing the Unlimited-plan sign-ups by 3.8%.

• Machine Learning Engineer

Memorial Sloan Kettering Cancer Center

Aug 2020–Feb 2021

New York, NY

- Designed and built a radiology-dedicated new language model RadioBERT by unsupervised pre-training of NLP model RoBERTa on radiology text data extracted from medical CT scan reports, and then fine-tuned the language model for multiple downstream tasks. Also developed guidelines for effective model verification.
- Orchestrated the state-of-the-art model for 700K+ reports (that achieved >96% accuracy) and elucidated the remaining team on the latest technologies in NLP, leading to a publication in the cancer research journal.
- This work enabled data engineers in the institution to transition radiology reports into lists of data items, each indicating which condition the patient is experiencing, reducing the look-up time from 40 to 3 mins per report.

• NLP Data Scientist (Intern)

Memorial Sloan Kettering Cancer Center

May 2020–Aug 2020

New York, NY

- Designed and implemented the complete NLP pipeline from scratch that predicts whether a metastatic (mets) disease of cancer is present in a given radiology report, using open-source models BERT and Longformer.
- Scaled the pipeline addressing 1 mets task to 13 mets tasks (such as kidney mets, liver mets, lung mets, spleen mets, etc. among others), and from 1K reports to 700K reports, boosting its usability and adoption rate by 80%. Also implemented the optimization job for hyper-parameter tuning.

- **Algorithm Engineer** (Intern) Jun 2019–Sep 2019
KLA–Tencor Corp. Milpitas, CA
 - Proposed and prototyped a no-reference IQ estimation technique which involved: extracting patches of images, extracting features from these patches, and then modeling these features using a multivariate gaussian model.
 - Implemented the Image Quality (IQ) metrics: signal-to-noise ratio, contrast-to-noise ratio, sharpness and rotation in each image (in Python) which provided accurate IQ scores for 94% of the given images, and is 4x faster than the prevalent Matlab pipeline.
- **Research Intern - Machine Learning** Jul 2017–Apr 2018
Tata Institute of Fundamental Research (Homi Bhabha Centre For Science Education) Mumbai, India
 - Proposed and developed a Handwritten Character Recognition System for the script NaYaNa designed for dyslexic students, using stacked neural networks, along with a training-data generation system.
 - Bagged 1st prize for the system in DJ ASCII, the state-level project competition conducted in Mumbai (May'18)
 - Designed and built an Android application for native mobile users along with a web interface that increased the user satisfaction by 100%.

PUBLICATIONS

- **Prachi Rahurkar**, Matt Olson, Prasad Tadepalli
Human Adversarial QA: Did the Model Understand the Paragraph?
NeurIPS 2020 Workshop on Human And Model in the Loop Evaluation and Training Strategies
- Richard K. Do, **Prachi Rahurkar**^{*}, Lior Gazit^{*}, et al.
Patterns of metastatic disease in patients with cancer derived from natural language processing of structured CT radiology reports over a 10-year period
Journal of Radiology 301 (1), 115-122 (* responsible for NLP and Deep Learning work)

AWARDS

- 1st Prize at the State-level Project Competition **DJ ASCII** (*amongst over 909 participants*) 2018
- 2nd Prize at the Project Expo of Thadomal Shahani Engineering College (*amongst over 300 participants*) 2018

PROJECTS

- **YouTube AI Assistant**: Developed a web-app that answers questions about any YouTube video, using OpenAI's LLM agent. Implemented using LangChain and Streamlit.
- **Pets Name Generator**: Developed a web-app that suggests creative names for your pet based on the inputs, animal type and its skin color. Implemented using LangChain and Streamlit.
- **Question-Answering**: Conducted research to generate syntactically and semantically similar passages given a passage of text, in Question-Answering. For the generated passages that are successful in misdirecting language model's predictions, implemented linguistic categorization of adversarial attacks in reading-comprehension QA.

SKILLS

- **Languages**: Python, Java, Scala, JavaScript, TypeScript, Go, C, C++, Bash, Ruby, HTML, CSS, Rust, PHP
- **ML/AI tools**: PyTorch, Tensorflow, LLMs, NLP, Computer Vision, Deep Learning, Spark, Docker, Kubernetes
- **AWS tools**: EMR, InfraCDK, Lambda, S3, Redshift, CloudWatch, EC2, Kinesis, DynamoDB, SQS, CloudFormation
- **Database technologies**: Cassandra, MySQL, PostgreSQL, Redis, Firebase, MongoDB
- **Frameworks**: Git, Jira, Flask, Apache Kafka, React, Redux, Node.js, Express, Espresso, UNIX, GraphQL, Apollo, Angular, Spring, Vue, jQuery, JUnit, RabbitMQ, Jenkins, Django, Fast API