Rohan Chaudhury

(979) 985-8903 rohan.chaudhury@tamu.edu linkedin.com/in/chaudhury-rohan rohan-chaudhury-portfolio

Education

Texas A&M University, College Station, Texas **August 2021 – December 2023 (Expected)**
Master of Science in Computer Science, with Graduate Scholarship GPA: 4.0/4.0
Courses: Deep Learning, Software Engineering, Natural Language Processing, Analysis of Algorithms, Operating Systems, Data Mining

National Institute of Technology (NIT) Durgapur, India **August 2015 – May 2019**
Bachelor of Technology in Electronics and Communication Engineering Cumulative GPA: 9.25/10

Technical Skills

Languages:	Python, Golang, Java, C++, C, TypeScript, JavaScript, SQL, Ruby, C#, R, MATLAB
Cloud skills:	Amazon Web Services (Lambda, Dynamodb, S3, EC2, EMR, etc.), SAP HANA Cockpit, GCP, Azure, IBM Watson
Libraries & Tools:	Tensorflow, PyTorch, PySpark, Caffe, Keras, Hugging Face, React, Django, Flask, OpenCV, OpenAI, ONNX, Selenium, Artificial Intelligence Model Efficiency Toolkit, Snapdragon Neural Processing Engine SDK, Linux, Git, JIRA, Docker

Work Experience

Soft Interaction Lab | ***Graduate Research Assistant*** **Mar – May 2022 & Aug 2022 – Present**
Dept. of Visualization | *Python, C#, Tensorflow, PyTorch, Hugging Face, Unity* Texas A&M University, College Station

- Spearheaded the development of a Conversational Artificial Intelligence VR/Web application to serve as a virtual patient for TAMU School of Nursing students, replacing manual training methods. Adopted and highly acclaimed by the school
- Developed a conversational AI-driven interview training VR/Web application utilizing a refined gpt-3.5-turbo model, capable of resume analysis, context retention in custom built memory components, and personalized questioning.

TAMU InfoLab | ***Graduate Thesis Student*** **Aug 2022 – Present**
Dept. of Computer Science & Engineering | *Python, Tensorflow, PyTorch* Texas A&M University, College Station

- Conducting research on harnessing Dense Passage Retrieval, Retrieval Augmented Generation, & LLMs to advance question-answering performance on Multidoc2dial & Wizards of Wiki datasets under Prof. James Caverlee's guidance.
- Researching enhanced disfluency detection and removal with lightweight large language models (T5-base).

Amazon.com, Inc. | ***Software Development Engineer Intern*** **May 2022 – August 2022**
Java, JavaScript, TypeScript, AWS - Lambda, DynamoDB, Athena, S3, EC2, VPC, CDK, IAM Seattle, Washington

- Developed a full-stack software that procures run-time customer-data consumption details of internal services and analyzes it to show the data consumption statistics and access limitations for the individual services in a dashboard
- Enabled service owners to get a better perspective of the data utilization details, access limitations, and possible security breaches all in one place (with the help of this software), thereby saving 90% manual effort in finding them.

Qualcomm Technologies, Inc.| ***Associate Software Engineer*** **Nov 2019 – Aug 2021**
Artificial Intelligence Software Team | *SNPE, AIMET, Tensorflow, PyTorch, Hugging Face, ONNX* Hyderabad, India

- Optimized several trained Neural Network models (of Samsung, OnePlus, and other OEM customers) utilizing model compression, quantization and fine-tuning techniques, to run the models efficiently on DSP cores of Snapdragon chipsets
- Implemented critical feature requests in Snapdragon Neural Processing Engine SDK to enhance its functionalities
- Developed a new Recommendation System to give suggestions of similar Salesforce issues raised by customers in the past for newly raised customer issues, with a reported accuracy of 74% across various engineering divisions of Qualcomm
- Developed a widely used (more than 1000 internal users/month) Automation Software to automatically download (Selenium), intelligently parse, & generate error logs & reports from device crash dumps sent by customers in Salesforce
- Fixed critical Docker, bokeh server, and documentation bugs in AIMET (Artificial Intelligence Model Efficiency Toolkit)

Projects, Publictations & Presentations

Publications & Presentations | **2018–2023**

- Seo, J.H., **Chaudhury, R.**, Oh, J.H., Kicklighter, C., Arguello, T., Wells-Beede, E. and Weston, C., 2023, June. Development of Virtual Reality SBIRT Skill Training with Conversational AI in Nursing Education. In International Conference on Artificial Intelligence in Education (pp. 701-707). Cham: Springer Nature Switzerland. | *Paper presented virtually at the International Conference on Artificial Intelligence in Education 2023, Tokyo, Japan.*
- Garcia, B., **Chaudhury, R.**, Versaw, M., Back, J., Kwon, D., Kicklighter, Caleb., Taele, Paul., Seo, J.H., 2023, August. AllyChat: Developing a VR Conversational AI Agent Using Few-Shot Learning to Support Individuals with Intellectual Disabilities Accepted In Human-Computer Interaction - INTERACT 2023 - 19th IFIP TC 13 International Conference, York, UK, August 28 - September 1, 2023. | *Will present virtually at the INTERACT 2023 conference, York, UK*
- Mahata, S., **Chaudhury, R.**, Kar, R., Mandal, D. and Saha, S., 2018, July. Optimal Integer Order Approximation of Fractional Order Human Ear Simulator. In 2018 15th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON) (pp. 660-663). IEEE. | *Paper presented in-person at the ECTI-CON 2018 conference, in Chiang Rai, Thailand.*

Projects | *Python, C++, Java, PyTorch, Tensorflow, Numpy, Pandas, Javascript, Unity, Android Studio* **2019 – 2023**

- Built a web application by using few-shot learning on gpt-3.5-turbo to generate and post new blogs every hour Website
- Outperformed the baseline model in SemEval 2023 Task-6 for classifying Legal Documents based on their rhetorical roles by modifying baseline model with T5-large tokenizer and encoder & achieved an accuracy of 81.6% Github
- Developed Hierarchical Attention Network for Sentiment Analysis as described in the paper Hierarchical attention networks for document classification and achieved an accuracy of 86.25% using BERT embeddings as input Github
- Utilized GPT-2 text generations capabilities for sentiment analysis (on IMDB dataset) using both few-shot learning and fine-tuning. Obtained an accuracy of 90%. Visualized attention outputs to gain insights and improve accuracy. Github
- Designed an Adaboost classifier for face detection using Viola Jones algorithm with 97% accuracy. Medium article,Code
- Utilized (1) bayesian optimization for hyper-parameter tuning to train a custom Convolutional neural network and (2) fine-tuned pre-trained ResNet50 and MobileNetV2 models for Facial Expression Recognition (ICML 2013). Code
- Estimated public speaking anxiety from VerBio dataset using (1) FNN trained with features modified using filter and wrapper category selection methods and Principal Component Analysis, (2) RNN, GRU, LSTM networks. Poster, Code
- Implemented efficient collaborative filtering and SVD++ matrix factorization as described in Korenś 2008 paper "Factorization meets the neighborhood: a multifaceted collaborative filtering model". Colab, Github
- Discovered and plotted interesting associations by analyzing US congress tweets dataset using word2vec, hugging face models, t-SNE, PCA, and k-means. Colab, Github. Initially explored the dataset using PySpark and graphframes. Code

Awards & Achievements

1. Awarded funding from the Academy of Visual and Performing Arts (AVPA) of Texas A&M University to present my research paper at the 24th International Conference on Artificial Intelligence in Education, AIED 2023.
2. Received Scholarship of $10,205/year for 2 years in a row from the Department of Computer Science and Engineering of Texas A&M University, College Station
3. Earned multiple Professional Excellence Awards from Qualcomm for independently developing two pivotal Natural Language Processing and automation-based software, and for enriching the Artificial Intelligence Model Efficiency toolkit (AIMET) through significant open-source contributions, all while exceeding the scope of my core responsibilities.

Trainings & Certifications

2023	Data Parallelism: How to Train Deep Learning Models on Multiple GPUs	NVIDIA Deep Learning Institute
2018	Machine Learning, Online course authorized by Stanford University	Coursera
2018	Neural Networks and Deep Learning, Course authorized by deeplearning.ai	Coursera
2018	Applied AI with Deep Learning,Course authorized by IBM, Badge Link	Coursera
2018	Google Cloud Platform Big Data and Machine Learning Fundamentals	Coursera