Website: https://rohansleekha.github.io/ **Open to relocate**

Rohan Singh Leekha

(Active Secret Clearance DoD)

Email: rohansl@vt.edu **Cell:** +1-540-824-8637

EDUCATION

- **M.S. Computer Science** (Virginia Tech Blacksburg) Fall 2021 – May 2023 *(GPA – 3.85/4)*

TECHNOLOGY STACK

- **Skills:** Natural Language Processing (NLP), Computer Vision (CV), Deep Learning, Machine Learning, Feature Engineering, GitHub, speech recognition, audio classification, language modeling, ASR, Ensembling, object detection and model stacking.
- **Big Data**: GCP, Kubernetes, AWS, Database Management, Docker, Hadoop, Apache, Mongo DB, Scala, MLOps, MapReduce, ELK stack.
- **Languages:** C++, Python, R, SQL, PowerShell, Bash.
- **Libraries:** Scikit-learn, NumPy, SciPy, pandas, Keras, TensorFlow, PyCaret, NLTK, RegEx, FastAI, PyTorch, Huggingface
- **Algorithms:** CNN, BERT, LSTM, RNN, XGBoost, Random Forest, Decision Tree, GAN, ANN, Q-learning, CatBoost, Wav2Vec, Ensembling, Blending and model stacking, YOLO.

WORK EXPERIENCE

Massachusetts Institute of Technology (MIT) Lincoln Laboratory (Group 52) (May 2022 –Aug 2022)

- Authorship verification and style change detection using contextual word embedding and stylometric features. Achieved state-of-the-art performance on the PAN'22 dataset and reddit VeriDark dataset. *(Presented my findings at SCALE 2022)*
- Developed a novel Mixup text embedding technique to tackle largely imbalanced datasets.
- Contextual counter-narrative generation against online radicalization using CLIP and GPT-3.

Graduate Researcher - SAIL (Society + AI & Language) Virginia Tech (August 2021 – Present)

- Used BERT-based language model to analyse differences in linguistic patterns between politically distinct news networks. Used Integrated gradients to identify themes and phrases predictive of moral panic. (**Under Review at ICWSM)**.
- Conducting research work focussed on investigating communicative patterns of groupthink in political discussions around topics that are wrought with mis/disinformation and examining how racial bias in news media has evolved overtime using a combination of NLP techniques such as transformers and deep audio analytics.
- Built a multimodal model (Vision + Text) to analyze the spread of genocidal language in political memes.
- Adapted methodological intuitions from prior work to develop a new technique to quantify how semantic polarization – the semantic distance between how two entities contextually use an identical word – evolves over time in online social movements.

Data Science Intern – PERATON (Jan 2022 – May 2022)

- Predicting hospital readmission using ensembling algorithms and feature engineering.
- Mapping the spread of COVID-19 in at-risk communities using time-series datasets.
- Semantic mapping of feature columns using fine-tuned sentence transformers.

Lead Machine Learning Application Developer (Concentrix Corporation, June 2018 to July 2021)

- Improved internal speech recognition for user authentication system by using MFCC & probability-based GMM.
- Developed a friend recommendation system using collaborative filtering and link prediction
- Home-Agent monitoring system via movement tracking, background noise detection, shoulder surfing, and face verification.
- Developed a face hashing pipeline using VGG face and Face-Net model, to match face indexes. Accuracy 99%.

Publications

- Arora, V, **Leekha, R.S**, et al. "Transfer learning-based approach for detecting COVID-19 ailment in lung CT scan." *Computers in biology and medicine* (2021): 104575.
- Arora, V, **Leekha, R.S**, et al. "Facilitating user authorization from imbalanced data logs of credit cards using artificial intelligence." *Mobile Information Systems* 2020 (2020).
- Arora, V., Ng, E.Y.K., **Leekha, R.S**, et al. "Heart sound classification using machine learning and phonocardiogram." *Modern Physics Letters B* 33.26 (2019): 1950321.
- Arora, V., Ng, E.Y.K., **Leekha, R.S**, et al. "Health of things model for classifying human heart sound signals using co-occurrence matrix and spectrogram." *Journal of Mechanics in Medicine and Biology* 20.06 (2020): 2050040.
- Arora, V., Verma, K., **Leekha, R.S**, et al. "Transfer Learning Model to Indicate Heart Health Status Using Phonocardiogram." (2021).
- Arora, V., Mahla, S.K., **Leekha, R.S**, et al. "Intervention of Artificial Neural Network with an Improved Activation Function to Predict the Performance and Emission Characteristics of a Biogas Powered Dual Fuel Engine." *Electronics* 10.5 (2021): 584.
- Arora, V., **Leekha, R.S.** and Chana. "An Efficacy of Spectral Features with Boosted Decision Tree Algorithm for Automatic Heart Sound Classification." *Journal of Medical Imaging and Health Informatics* 11.2 (2021): 513-528.

PATENTS

- An IoT enabled EPCG device for nursing heart health distantly with artificial intelligence at cloud-based environment.
- Artificial intelligence-based system for conducting road safety audit using video streaming through IoT based devices.