Harshita **Diddee**

harshitadd@gmail.com linkedin.com/in/harshita-diddee github.com/harshitadd https://harshitadd.github.io

Education

2017 – 2021 **B.Tech, Computer Science and Engineering, Guru Gobind Singh Indraprastha University**
Major GPA - 9.5/10.0, CGPA - 9.46/10.0
Graduated as the Best Outgoing Student for the Class of 2021

Experience

Present – July 2021 | **SCAI Centre Fellow, SCAI, MSR, Bangalore**
- Developed Quantized and Distilled Machine Translation models (approximate size being <400MB) for extremelely low-resource languages (data <25K sentences).
- Evaluating the efficacy of using such lightweight models to provide dynamic assistance to data providers via assistive-translation interfaces.
- **Mentored by Kalika Bali,Principal Researcher**

Neural Machine Translation | Compression of Large-LMs

August 2022 – June 2022 | **Visting Pre-Doc for JSALT'22, CLSP, JHU, Baltimore, USA**
- Evaluated the generalizability of speech and text cross-lingual models to extremely low-resource languages.
- Exploring the development of mutlilingual variant of speechT5.
- **Participated in the Speech Translation for Under-Resourced Languages Track**

Neural Speech Translation | Low Resource Language Modelling

October 2020 – March 2021 | **Intern, AI4Bharat (IIT Madras), Remote**
- Implemented a tesseract-based OCR pipeline;
- Assisted the mining of parallel sentences from a dense (12M+), monolingual embedding space (between a target and source language) using FAISS
- **Mentored by Dr.Mitesh M. Khapra,Assistant Professor (CSE),IIT Madras**

Neural Machine Translation | Low Resource Language Modelling

March 2020 – November 2020 | **Research Intern, Indraprastha Institute of Information Technology (IIIT), Delhi**
- Validating a Cross-Silo Federated Learning (FL) hypothesis for heterogeneous clients having distributed and exclusive feature space.
- **Mentored by Dr.Koteswar Rao Jerripothula, Assistant Professor (CSE), IIIT Delhi**

Cross-Silo Federated Learning | Dropout Regularization

May 2019 – October 2019 | **Research Intern, Indian Institute of Technology, Delhi, Celestini Program India 2019 ,The Marconi Society**
- Developed a federated learning enabled custom deep learning model that powers VisionAir, an android application that predicts the Real-Time Air Quality Index of an image.
- Work published by TensorFlow
- **Mentored by Dr.Aakanksha Chowdhery, Google Brain**

Cross-Device Federated Learning | Computer Vision

Relevant Publications

June 2022 **The Six Conundrums of Building and Deploying Language Technologies for Social Good, ACM COMPASS, 2022, Authored by : Diddee et. al.**
I explored literature at the intersection of NLP and HCI to substantiate the importance of defining concrete tools that can assist Language Technologists in resolving non-trivial decision dilemmas during low resource, community-driven technologies.
ACM Digital Library

April 2021 **Samanantar : The Largest Publicly Available Parallel Corpora Collection for 11 Indic Languages, Transactions of ACL, Authored by : Ramesh et. al.**
I implemented a tesseract-based OCR pipeline; Additionally, I assisted the mining of parallel sentences from a dense (12M+), monolingual embedding space (between a target and source language) using FAISS.
Transactions of ACL

December 2020 **PsuedoProp at SemEval-2020 Task 11:Propaganda Span Detection using BERT-CRF and Ensemble Sentence Level Classifier, SemEval, 2020, Authored by Chauhan et. al.**
I implemented a sequential BERT-based model's inference pipeline. I also developed a post-processing pipeline that would bridge the structural difference between the outputs of the Sentence Level classifier and the fine-grained analysis BERT-CRF Model. We ranked 14th globally on Shared Task 11 of SemEval'2020 collocated with COLING.
ACL Anthology

Honors and Grants

May 2022 Represented our team's work at the Panel for ComputeEL-5, ACL; **ACL 2022**
March 2021 Received the Directors' Scholarship to support my Master's Education at the Carnegie Mellon University [Admission Withdrawn]**Carnegie Mellon University**
September 2020 Received a grant to present my work at The 35th IEEE/ACM International Conference on ASE by **ACM**
August 2020 Selected to attend the Google AI Summer School 2020 by **Google AI**
March 2020 Received a $1000 USD Travel Grant to attend the TensorFlow Dev Summit at Sunnyvale,CA by **Google LLC**
October 2019 Runner's Up at the Celestini Prize India 2019; Granted by **The Marconi Society, Google**
October 2019 1st Runner's Up at the Singapore India Hackathon; Awarded by the **Prime Minister Of India and the Former Education Minister of Singapore, Government of India, Government of Singapore**
March 2019 Ranked 1st Nationally at the e-Yantra National Finals; Awarded by the **MHRD (Government of India)**

Relevant Projects

CodeFed : Federated Learning Enabled Code-Switching
September 2020 - June 2021

Under Review
I am contributing to the development of a federated learning-trained code-switching detection model that can be deployed to collaboratively trained be leveraging private data from multiple users, without compromising on their privacy. After validating the paradigm's viability for the chosen use-case, a label generation system will be devised to allow real-time on-device training of the models on the client' device.

Code-Switching in Audio | PyTorch | Speech Modelling

Native Accent Sensitive Voice Cloning
January 2020 - July 2020

I contributed to the architecture development of a native-accent sensitive voice cloning system, and implemented the accent identification encoder of the model; It's main use would be to model realistic voice to be used by education platforms where a seed audio sample of a native speaker could be used to dub the subtitles of a MOOC in the user's indigenous language.

Speech Modelling | PyTorch