Sree Bhargavi Balija

858-319-6721 sbalija@ucsd.edu linkedin.com github.com portfolio Location: San Diego, CA

Education

University of California San Diego **March 2024**
Master of Science in Machine learning and Data science ***CGPA: 3.5/4.0***
Indian institute of Technology Hyderabad **July 2020**
Bachelor of Technology in engineering ***CGPA: 9.1/10***

Technical Skills

Languages: C/C++, Python, Java, Javascript, Angular, Kotlin, Prolog, Perl
Web Technologies/Frameworks: Google cloud, Apache Spark, Hadoop, Docker, Firebase, Hugging Face, Flower
Databases: Oracle SQL, MySQL
Data Science: Bert language models, Classical ML, DL, NLP, Explainable AI, Federated learning, Computer Vision,

Relevant Coursework

- Statistical learning
- NLP
- Search and optimization
- Learning Algorithms
- Artificial Intelligence
- Deep generative models
- Recommendation systems

Job Experiences

ServiceNow, Software engineer | *Java, Js, Angular, Eclipse, Github* **June 2020 – August 2022**

- Worked on integrating multiple rest api's with ITSM workflows for adding capabilities like **Citrix cloud virtual systems access** and Request item flow to the **Virtual bot**, further developed the topics **Natural language understanding** models for **Intelligent conversation flows**
- Implemented **automated testing strategies** encompassing unit tests, integration tests, functional tests, and performance tests to ensure code quality and reliability.

Research Intern, Shiley Eye Institute | *Matlab, Python, C++* **Dec 2023 - Ongoing**

- Working on **deep learning models** for the characterization of the **optic disc phenotypes** in glaucoma patients

Academic Projects & Research Experience

Federated fine tuning of heterogeneous Large Language Models | *Python* **Dec 2023**

- Developed a novel **Federated LLM** technique from finetuned **BIOBERT** models of multiple decentralized nodes (**Edge devices**), each local model has been finetuned on their own local device data
- Implemented a bot by fine-tuning **NLP** queries on the **LLaMA** (Language Model for Many Applications) model which involves a structured approach that combines understanding the model architecture, preparing the dataset for fine-tuning and then deploying the bot on website
- This framework addresses the **privacy, data scarcity issues** and specifically applicable for **NLP tasks**
- Building a search engine using **Falcon LLM** specifically for academic and research papers that understands the context and semantics of **user queries**, providing more relevant and precise results, and even summarizing research findings.

Federated learning clients side pruning through mixed precision quantization techniques **Sep 2023 - Ongoing**

- Working on novel client sided mixed precision quantization technique which out performs the **Hessian awareness spectrum quantization technique** in terms of inference speed
- Developed new client pruning method using **conformal predictions** which selects the most efficient clients for high global model performance.
- Working on Novel **Interpretable federated learning** method using additive models and shapley values.

Apache-spark-structured-streaming | *Python* [code] **March 2023**

- Developed an end-to-end architecture leveraging **Twitter API** and Python module Tweepy to stream and perform sentiment analysis on Twitter data, resulting in the creation of a Kafka topic for real-time data analysis.
- Utilized **PySpark SQL** to calculate the average sentiment score of streamed tweets and implemented a function to categorize tweets as positive, negative, or neutral based on sentiment scores and insights extraction capabilities.

Accolades/ Online Certifications

- **Academic excellence award**, IIT Hyderabad **2018**
- UCSD ECE **Summer research internship** scholar, UCSD **2023**
- **Teaching Assistant**, Introductory courses in physics and chemistry departments website management **2018**
- Silver medal, International Master **Mathematics Olympiad** **2013**
- Representive of IITH in social online innovation collaborative **hackathon** **2020**
- Achieved Skill development incentive program award, **ServiceNow** **2021**

Publications

- Developing Communication efficient asynchronous peer to peer federated LLMs with blockchain AAAI, Stanford Uni