SAN DIEGO, US • SBALIJA@UCSD.EDU • 858-319-6721

SREE BHARGAVI BALIJA

Machine Learning and Data Science Professional

EMPLOYMENT HISTORY

RESEARCHER **Apr 2024**
US Meta Research Team

- Spearheaded CHAI paper enhancement through hybrid sparse attention mechanisms integration and targeted fine-tuning.
- Pioneered memory reduction strategies, achieving 70% efficiency in transformer models.
- Optimized large-scale NLP tasks through advanced clustering techniques

CONTRIBUTOR **Nov 2024 - Present**
ML Commons Open Source Platform

- Led the development of multimodal projects, translating video to text
- Pioneered transformer-based models, enhancing automated video understanding and summarization
- Advanced open-source AI, driving efficiency in multimodal learning systems

ARTIFICIAL INTELLIGENCE INTERN **Apr 2024 - Jun 2024**
Radical AI

- Led AI application development using top-notch frameworks like Langchain, OpenAI, Google Gemini.
- Engineered multi-modal interactive elements and AI-powered games, fostering an engaging learning environment.
- Effectively utilized machine learning and data science skills to drive project success.

SOFTWARE ENGINEER **Jun 2020 - Aug 2022**
ServiceNow

- Led the integration of multiple REST APIs with ITSM workflows, enhancing capabilities such as Citrix cloud virtual systems access and virtual bot item request flows.
- Created a comprehensive dashboard, providing pre-built analytics for eight key metrics including customer satisfaction and cost savings.
- Developed a Java API for managing compaction statistics, enabling both periodic and user-triggered compaction and job cancellation.

EDUCATION

MASTER OF SCIENCE IN MACHINE LEARNING AND DATA SCIENCE **Jun 2024**
University of California San Diego

BACHELORS OF TECHNOLOGY IN ENGINEERING **Jul 2020**
Indian Institute of Technology Hyderabad

SKILLS

Python *(Expert)*, Java, JavaScript, Angular, C++, R, Fortran, Prolog, Perl, Kotlin, Swift, PyTorch, Kubernetes, NLTK, NLP, Machine Learning, Deep Learning, Federated Learning, SQL, Django, ETL, Azure, GCP, Elasticsearch.

ADDITIONAL INFORMATION

PUBLICATIONS

- Building Communication Efficient Peer-to-Peer Federated LLMs with Blockchain, AAAI, Stanford University (link)
- FedNAM: Performing interpretability analysis in federated learning context, ICLR 2025 (In review)
- CPTQuant - A Novel Mixed Precision Quantization Techniques for Large Language Models (link)
- FedNAM+: Executing Interpretability Analysis using Novel Conformal Predictions method, CVPR 2025 (In review)

AWARDS AND CERTIFICATIONS

- Skill development incentive program award, ServiceNow 2021
- Academic excellence award, IIT Hyderabad 2018
- Silver medal, International Master Mathematics Olympiad 2013
- Building LLMs Efficiently, NeurIPS 2024 Challenge (Ongoing)