Mihir Mangesh Pavuskar

(213) 691-9326 | pavuskar@usc.edu | linkedin.com/in/mihir-pavuskar-658337139 | mihirp.me

EDUCATION

University of Southern California **Los Angeles, CA**
Master of Science in Computer Science GPA: 3.95 | Dec. 2024
Relevant Courses: Artificial Intelligence, Algorithms, Machine Learning, Natural Language Processing, Deep Learning, Perception

Vellore Institute of Technology **Vellore, India**
Bachelor of Technology in Computer Science and Engineering GPA: 3.82 | Aug. 2022

SKILLS

Machine Learning: Tensorflow, PyTorch, NLP, Computer Vision, Algorithms, SciKit Learn, LangChain, MATLAB, Python, C++
Data Science: Data Structures, SQL, DynamoDB, MongoDB, PostgresDB, Hadoop, Spark, R, Python, Redis, Visualization, Pandas,
Full Stack: Reactjs, Nodejs, NextJS, Docker, AWS, REST APIs, CI/CD, Git, Go, Typescript, Javascript, HTML, CSS, WebAssembly

PROFESSIONAL EXPERIENCE

AI Software Developer Intern **Los Angeles, USA**
Tikr Media Aug. 2024 – present

- Develop and **fine-tune LLM and RAG** systems using **PyTorch** for automated campaign content generation, resulting in 25% increase in user acquisition and streamlined marketing operations.
- Implemented sophisticated **PII redaction mechanisms** and **synthetic data generation** for LLM training and inference, ensuring GDPR compliance while maintaining model performance across 10,000+ user records

Software Developer, Frontend **Bengaluru, India**
MURF AI Sept. 2022 – Dec. 2022

- Spearheaded migration of legacy video/audio editing studio to modern React architecture using Redux and TypeScript.
- Analyzed **Google Analytics** logs to identify and resolve bugs and **devise fail-safes, lowering crash rates by 40%**.
- Optimized studio for data-intensive projects by applying **UI virtualization** and modifying data flow across **ReactJS** components and **Redux**, making app function 5 times faster.

Natural Language Processing Researcher **Geelong, Australia**
Deakin University Feb. 2022 – July 2022

- Architected end-to-end **financial question-answering system** using ensemble of **BERT-based models (TaPaS, ALBERT, HybridR)**, achieving state-of-the-art accuracy of 72% on **FinQA benchmark dataset.**
- Engineered **custom data preprocessing methodology** integrating **structured and unstructured financial data** processing over 1,000 financial documents with 99.7% accuracy

Student Research Intern **Bengaluru, India**
Samsung Dec. 2019 – May 2020

- Designed **intelligent text normalization system** for **ASR** using ensemble of **BERT and XGBoost models**, improving model accuracy by 4.4% on production data.
- Engineered **end-to-end data pipeline** for **slang normalization**, with custom preprocessing algorithms for handling non-standard words and internet slang, achieving 98.75% classification accuracy across 17 test batches of 500,000 samples each

PROJECTS

CALM-NAV: Confidence-Adaptive Learning and Monitoring for Vision-Language Fall 2024

- Led development of **Adaptive Human-in-the-Loop Vision-Language Navigation** system integrating confidence estimation with **VLN-BERT**, achieving 62.2% success rate on **R2R validation** seen and 52.8% on unseen environments.
- Developed **confidence calibration pipeline** using **Monte Carlo Dropout** and **entropy measures**, reducing Expected Calibration Error from 12% to 8% while maintaining real-time performance.

Mesh2Splat: Deep Learning method to convert Meshes to Gaussian Splats Spring 2024

- Spearheaded design of **end-to-end deep learning pipeline** leveraging **PointNet++ architecture** for converting **3D meshes** to **Gaussian splats**, achieving 93% accuracy on test data.
- Created Objaverse-Splats, the first **comprehensive dataset** of graphics shapes paired with Gaussian splat representations, enabling training of robust ML models for 3D object conversion.

MorphVLM: Fine-tuning Large Multimodal Language Models on niche Domains Fall 2023

- Architected novel **fine-tuning methodology** for **Vision Language Models** on niche domains by substituting language components with **pre-trained LLMs**, achieving 45.5% accuracy on **VQAv2 benchmark** while accelerating convergence 3x
- Utilized **Low-Rank Adaptation (LoRA)** to reduce trainable parameters by 30% to 11.4M further cutting training time by 40%

Online Pente Playing AI Agent Spring 2023

- Implemented leveraging **alpha-beta pruning algorithm** running at depth 5 with **C++** capable of defeating **Random, Minimax and Level 1 agent** on pente.org.
- Improved performance by reducing search space and through optimizations like **Forward Pruning, Move Ordering**, etc.
- Devised memory efficient agent and compiled to **web assembly** to produce near-native performance on client side and **deployed** as a user-friendly **ReactJS** app using **Netlify**.