Pavan Kumar Anand

+1 206-387-9639 | pavanka1547@gmail.com | https://www.linkedin.com/in/pka2000/ | Seattle, WA

Education

University of Washington (UW), Seattle, WA **Sep 2022 – Jun 2024**

Master of Science (MS) in Computer Engineering (Major: Data Science)

Relevant Courses: Machine Learning, Computer Vision, Natural Language Processing (Audit), AI for Engineers, Introduction to Statistical Learning, Introduction to Database Systems, Data Visualization

Indian Institute of Information Technology (IIIT), Sri City, Chittoor, India **Aug 2017 – Jun 2021**

Bachelor of Technology (BTech) in ECE

Relevant Courses: Data Structures and Algorithms, Probability Theory and Statistics (M3), Pattern Recognition, Statistical Data Analysis, Advanced Statistical Methods, Applied Software Engineering, Computer Architecture and Operating Systems

Publications

- Quilt-1M: One Million Image-Text Pairs for Histopathology. **NeurIPS ORAL 2023**. Paper Link
- Random Fourier Features based Post-Distortion for Massive-MIMO VLC. **IEEE ICCSPA 2020. [First Author]** Paper Link

Skills

Languages/Databases: Python, R, C/C++, JavaScript, SQL

Tools/Software: AWS, Azure, MySQL, Django, Flutter, Git/GitHub, HTML, CSS, ReactJS, NodeJS, MongoDB (NoSQL), Tableau, PowerBI, MS Office, LaTeX, Mathematica.

Machine Learning/Data: PyTorch, OpenCV, Keras, Generative AI, Large Language Models (LLMs), Computer Vision, Deep Learning, Natural Language Processing, Machine Learning Algorithms, Time Series Forecasting, A/B Testing, Statistical Tests.

Work Experience

Icertis | Data Science and Analytics Intern **Jun 2023 – Aug 2023**

- Extracted and analyzed over **300K** support tickets from ServiceNow, identifying key issues to refine training content.
- Developed and implemented an automated ticket classification system using **SVM, XGBoost, BERT** and **RoBERTa** to classify a substantial volume of customer support tickets into 12 categories achieving a reduction of manual effort by **95%** in time.
- Reduced **operational costs** compared to alternative models like **GPT-3.5**, ensuring sustainable and budget-conscious deployment.

NanoString Technologies Inc. | Machine Learning Engineer (Capstone) **Jan 2023 – Jun 2023**

- Trained a **UNet**-based DECODE model in **PyTorch** to achieve **40%** improvement in localization compared to traditional image processing methods.
- Employed **transfer learning** to expedite the identification of RNA molecules across four color emission PSFs, resulting in reduced training time and computational resources.
- Translated the model in PyTorch to C# using **ONNX** to deploy the model into NanoString's Imaging product, CosMx. **Won 2nd place** among **53** industry mentored projects in UW ECE 2023 Capstone Showcase Event.

UW Graphics and Imaging Lab | Machine Learning Researcher **Jan 2023 – Jun 2023**

- Worked towards curating the largest vision-language histopathology dataset Quilt-1M containing **768,826 image and text pairs** comprising **200k** samples using a mixture of models, including **LLMs**, handcrafted algorithms, and automatic speech recognition.
- Demonstrated the value of Quilt-1M by **fine-tuning** a **pre-trained CLIP** model utilizing the **ViT-B/32** architecture with **GPT/77**. The model **outperforms** state-of-the-art **foundation models** on both zero-shot and linear probing tasks for classifying new histopathology images.

Gyrus AI | Machine Learning (AI) Engineer **Sep 2021 – Jun 2022**

- Pioneered an extensive synthetic dataset to enhance superresolution capabilities of low-quality images and videos, collaboratively evaluating models like **EGVSR** and **TecoGAN** using computer vision-based algorithms in **OpenCV** for optimal model selection.
- Conducted in-depth research on video compression algorithms and codecs such as **H.264** and **HEVC**, integrating these insights to develop algorithms that improve video quality while maintaining efficient compression, enhancing storage and transmission solutions for clients.
- Applied **optical flow** techniques to improve the temporal consistency of super resolved video frames, ensuring smooth transitions and higher visual fidelity.

Projects

Deployment of Fine-Tuned LLMs on Edge Devices for Medical Applications **Jun 2024 – Present**

- Fine-tuned state-of-the-art LLMs such as **LLaMA 2 7B**, **LLaMA 3 8B**, and **Gemma 2B** using **PEFT** techniques like **LoRA** and **QLoRA**; reduced memory footprint via **pruning**, **quantization**, and **knowledge distillation**, achieving a **75%** reduction in model size with less than **4%** performance loss; experimented with **2, 3, 4**, and **8-bit** quantization; fine-tuned on a medical institution tuning dataset with **10k** prompt completion pairs for deployment on edge devices.

Evaluating Fine-Tuning Techniques for Diffusion Models **Mar 2024 – Jun 2024**

- Evaluated **LoRA**, **Textual Inversion (TI)**, and **DreamBooth** on Stable Diffusion, reducing computational overhead and enhancing visual concept integration. Achieved **15%** improvement in **FID** scores with DreamBooth, demonstrating high-quality, contextually relevant image generation.

Responsibilities

- **Graduate Teaching Assistant** at UW's Paul G. Allen School of Computer Science and Engineering for courses **CSE416: Introduction to Machine Learning** and **CSE160: Data Programming**.