Subha Vadlamannati

(206)-258-9108 | subha2@stanford.edu | linkedin.com/in/subhavadlamannati | github.com/subha-v

Education

Stanford University GPA: 3.8/4.0 — Palo Alto, CA

Bachelor of Science in Computer Science — *June 2028*

Coursework: Machine Learning (CS229), Reinforcement Learning (CS234), Artificial Intelligence (CS221), Continuous Mathematical Methods for Machine Learning (CS205L), Design & Analysis of Algorithms (CS161), Operating Systems Principles (CS111), Probability for Computer Scientists (CS109), Mathematical Foundations of Computation (CS103), Computer Graphics (CS148), Linear Algebra, Differential Equations (Math 51, 53).

Publications

Vadlamannati, S., Pasumarthi, A, Yang, D., Neubig, G., Khanuja, S., et al. (2025). HILITE: Human-in-the-loop Interactive Tool for Image Editing. ***Proc. 2025 Conference on Empirical Methods in Natural Language Processing (EMNLP) HCI + NLP.***

Vadlamannati, S. and Solgi, R. (2023). Partial Tensorized Transformers for Natural Language Processing. ***Proc. 16th International Conference on Agents and Artificial Intelligence.***

Vadlamannati, S. and Gozde, S (2023). Metric-Based In-context Learning: A Case Study in Text Simplification. ***Proc. 16th International Natural Language Generation Conference.*** *Nominated for Best Paper.*

Experience

Machine Learning Engineer — June 2025 – September 2025

VALUENEX — *Palo Alto, CA*

- Engineered a high-precision indoor positioning system using Apple's Nearby Interaction (UWB) framework, achieving centimeter-level distance accuracy (±10cm) across a distributed mesh network of iOS devices.
- Developed a multi-anchor mesh architecture with dual-stack IPv4/IPv6 support, leveraging Bonjour/mDNS for zero-configuration discovery and real-time distance tracking across multiple concurrent nodes.

Research Intern — June 2024 – January 2025

Stanford Artificial Intelligence Laboratory (SAIL) — *Stanford, CA*

- Worked with Stanford's SALT lab and CMU's Neulab to develop HILITE, an open-source interactive image-editing platform that integrates six state-of-the-art diffusion models (InstructPix2Pix, AnyDoor, etc.) to generalize across diverse editing tasks, including style transfer and object swapping.
- Implemented a human-in-the-loop workflow using Next.js, FastAPI, and RunPod serverless GPUs to iteratively collect user feedback and specific hyperparameters, creating a parallel dataset for future VLM fine-tuning.

CEO, Founding Fullstack Engineer — March 2021 – Present

OpenNLP Labs (formerly Linguistics Justice League) — *Seattle, WA*

- Developed 3 fullstack web applications (Edulang, Polyglo, HeritageHub) alongside CMU's Neulab & Stanford SAIL used by over 1000+ refugees for resources in 108+ languages. Collected user feedback & fine-tuned LLMs.
- Secured $185k+ in funding from Microsoft, T-Mobile, etc. Featured on NASDAQ's billboard on Times Square.

Research Intern — June 2023 – October 2023

University of California, Santa Barbara — *Santa Barbara, CA*

- 1 of 77 out of 4000 applicants selected to participate in the Research Mentorship Program (RMP). Developed a novel PTNN (Partially Tensorized Transformers) approach to compress vision-language models (BERT, ViT) by 53%, improving accuracy by up to 5% without post-training adjustments. Presented & published findings at MIT's Undergraduate Research Technology Conference & GRITx.

Projects

Multi-resolution Satellite Fusion for Canopy Height Prediction | *Python, PyTorch, CUDA, AWS* — Dec 2025

- Developed a multi-resolution fusion CNN (3.6M parameters) to predict global forest canopy height from optical satellite imagery
- Architected distributed training infrastructure across 8 GPUs using PyTorch DDP with NCCL backend, implementing GPU-resident data loading (14GB dataset on-GPU), mixed precision training, and automatic batch scaling with linear learning rate rules.
- Built end-to-end data curation pipeline processing 3TB+ of raw satellite imagery from AWS, implementing quality filtering on GEDI LiDAR measurements, forest masking via ESA WorldCover, and spatial GroupKFold splitting to prevent data leakage.