Kesavan Ramakrishnan

916-990-8055 | kesavanr@stanford.edu | linkedin.com/in/kesavan-ramakrishnan | github.com/kesavanramakrishnan

Education

Stanford University — Stanford, CA
Bachelor of Science in Computer Science, Grade: 4.01/4.00 — *Expected June 2027*

Experience

AI/ML Systems Research Intern — June 2025 – Present
AMD Research & Development — *San Jose, CA*

- Implemented a Stream-K style, work-centric backward pass for attention, inspired by Lean Attention, to resolve work-imbalance and improve performance for long-sequence LLMs on current and future Instinct GPUs.
- Optimized the Lean Attention forward pass to significantly boost GPU throughput for long-sequence models.
- Enhanced memory access patterns by implementing advanced kernel techniques, including memory swizzling.
- Tuned kernel launch parameters for the CDNA architecture to maximize occupancy and execution efficiency.

ML Systems Researcher — March 2025 – Present
Stanford Artificial Intelligence Lab: Hazy Research Group — *Stanford, CA*

- Ported ThunderKittens, tile-based DSL from NVIDIA to AMD GPUs by adapting CUDA abstractions to HIP.
- Wrote high-performance HIP kernels using TK's tiled abstractions, for simple ops (Vector Add, GEMM, etc.).
- Utilized profiling tools for performance debugging (e.g., SMEM bank conflicts); applied swizzling techniques to restructure memory layouts and improve memory access patterns, improved throughput on various kernels.

Computer Science Teaching Assistant — December 2024 – Present
Stanford University: CS 198 — *Stanford, CA*

- Teaching assistant for Stanford's largest introductory programming class in Python *(CS 106A/106B)*.
- Hosted weekly sections for students and debugged students' programs during weekly office hours (LaIR).
- Graded weekly assignments, held interactive feedback sessions w/ students to provide a better learning experience.

Web Development Intern — June 2023 – July 2023
California Department of Technology — *Sacramento, CA*

- Worked under Data and Geospatial services + Web Development teams on developing state website services.
- Utilized frameworks such as OpenStreetMap and Leaflet, and mapping softwares, including ESRI and ArcGIS.

Projects

Accelerated MRI Reconstruction with SwinUNet | *PyTorch, FastMRI* — March 2025 – June 2025

- Designed a hybrid SwinUNet architecture that fuses a U-Net encoder-decoder with shifted-window attention transformers, enabling accurate reconstruction of undersampled knee MRIs from the fastMRI dataset.
- Outperformed baseline UNet, raised PSNR to 33.1 dB and SSIM to 0.727 while preserving fine anatomical details.
- Optimized training, final model delivers 43 ms inference per slice—fast enough for clinical workflows.

TreeCycle | *Swift, Vision Pro, FastAPI, Brev.dev, Python* — Feb. 2025 – May 2025

- Built VR application on Apple's Vision Pro to educate on sustainability for waste management.
- Designed iOS application that uses YOLO model trained on Brev.dev platform to classify waste from an image.
- Game is simulated with classified waste, environment dynamically changes based on decisions made by user.

Scollioexercise | *Inventor / Lead Programmer* — June 2020 – Aug. 2024

- Led a team to create a medical device that supports people with scoliosis during exercise or physical activity.
- Raised $20K+ in funding from investors for research and development of a viable prototype for launch.
- Device detects pressure or weight on the spine using point watchers placed strategically around vest.
- Presented to doctors and Disney engineers across the world as 1 of 20 teams at 2021 Global Innovation Awards.

Technical Skills

Languages: Java, Python, C, C++, JavaScript, HTML/CSS, Swift
Libraries: PyTorch, NumPy, React, Flask, Triton, CUDA, ROCm
Software Tools: Git, AutoCAD, Blender, MATLAB, Microsoft Azure, AWS