**Ravi Ghadia**

rghadia@utexas.edu

ghadiaravi13.github.io

15-Sept-2024

I am writing to express my strong interest in joining [COMPANY] as a [POSITION], where I can bring my expertise in **Machine Learning**, **Deep Learning Optimizations**, and **Computer Architecture** to contribute to your innovative projects. I hold a Master’s degree in Electrical and Computer Engineering from the University of Texas at Austin, where I focused on Computer Architecture and Embedded Systems, and I have gained valuable industry experience at **NVIDIA** as a GPU Architect.

During my time at NVIDIA, I was actively involved in developing high-fidelity statistical models to diagnose inefficiencies in GPU architecture. I applied **Reinforcement Learning (RL)** and **Convolutional Neural Networks (CNNs)** to solve optimization problems, such as maximizing performance per watt (MaxQ) and enhancing bin optimization algorithms. My experience extends to using **predictive machine learning models** for workload performance projections in HPC and large language models (LLMs), as well as building Monte Carlo simulation environments to guide strategic product roadmaps for datacenter systems. This combination of practical experience in machine learning and deep understanding of hardware-software interactions has shaped my approach to developing scalable, efficient systems.

I am deeply passionate about continuing my work at the intersection of **ML/DL optimization** and **Computer Architecture**. My recent research at the University of Texas focuses on tuning the **KV cache size** to optimize inter-token generation latency during **LLM inference**, demonstrating my commitment to advancing deep learning performance. I am eager to apply my technical knowledge in large-scale machine learning systems, GPU architecture, and deep learning optimization techniques to [COMPANY]’s cutting-edge projects. My goal is to push the boundaries of efficiency in AI and computing systems, making high-performance machine learning more accessible and sustainable.