Prajwal Sallapalli

San Francisco, CA | prajwal.sallapalli2@gmail.com | 341-800-1435 | LinkedIn | GitHub

PROFESSIONAL SUMMARY

Software Developer and ML Research Scientist with experience building high-performance networking software, distributed systems, and AI/ML systems. Proven track record in C++, Python, and Linux, with deep expertise in multithreading, concurrency, large language models, neural network architectures, diffusion models, and performance optimization. Experienced developing production software, conducting applied AI research, and designing scalable infrastructure with a focus on reliability, efficiency, and real-world impact.

EXPERIENCE

ML Research Scientist May 2026 – Present

Roogen AI

- Developed and evaluated novel LLM architectures, improving benchmark accuracy by 10% through architectural modifications, training optimization, and model scaling experiments.
- Designed custom neural network architectures in PyTorch, reducing training time by 15% through optimized computation graphs, parallelization strategies, and efficient data pipelines.
- Researched and optimized diffusion-based generative models, reducing GPU power consumption by 5% while maintaining generation quality and inference performance.
- Built end-to-end training, evaluation, and benchmarking frameworks supporting rapid experimentation across multiple model architectures and datasets.
- Implemented fine-tuning, transfer learning, and parameter-efficient adaptation techniques to improve model performance on domain-specific tasks.
- Performed large-scale hyperparameter optimization and model profiling to enhance convergence stability, inference efficiency, and overall system scalability.
- Partnered with research and engineering teams to deploy and operationalize machine learning solutions in production environments.

Software Developer Aug 2023 – Jan 2025

Hewlett Packard Enterprise (HPE)

- Optimized Linux firewall packet-processing systems (DPI, AppRF), improving throughput by 15-20% in production networking environments.
- Developed C++ automation and validation frameworks for router and switch firmware testing, improving software reliability by 30%.
- Investigated complex software failures using debugging, profiling, runtime analysis, and root-cause investigation across networking and Linux-based systems.
- Built automated testing and validation workflows supporting large-scale production deployments.
- Collaborated with engineering teams to improve scalability, reliability, and operational performance of packet-processing pipelines and routing infrastructure.

Software Developer Intern Jan 2023 – Aug 2023

Hewlett Packard Enterprise (HPE)

- Developed automation tools and validation workflows for networking software using Linux and C++.
- Identified and diagnosed software failures and performance bottlenecks using Linux debugging tools.

PROJECTS

Low Latency Order Book Engine
https://github.com/Sallapalli-Prajwal/low-latency-orderbook

- Built a high-performance multi-threaded C++ matching engine implementing price-time priority with sub-ms latency.
- Processed 10K+ orders/sec using lock-efficient concurrent data structures and optimized execution paths.
- Designed order matching logic, market order processing, and performance-critical system components.
- Applied concurrency techniques, memory optimization strategies, and low-latency design principles.
- Benchmarked throughput, latency, and scalability under varying workloads.

Distributed Data Processing Framework https://github.com/Sallapalli-Prajwal

- Built a distributed processing framework supporting concurrent task execution and workload balancing.
- Implemented scheduling, monitoring, runtime analytics, and performance measurement systems.
- Designed scalable workflows inspired by cloud-native distributed systems.
- Collected runtime metrics and analyzed performance behavior under varying workloads.

GPT Implementation & Instruction Fine-Tuning
https://github.com/Sallapalli-Prajwal/LLM

- Implemented GPT-style transformer architecture in PyTorch.
- Built training, evaluation, and inference workflows while analyzing model performance and scalability.

AI Resume Builder (MERN + LLM) https://github.com/Sallapalli-Prajwal/resume-builder

- Built and deployed an AI-powered application used by 500+ users.
- Developed scalable backend APIs and cloud-hosted services supporting production workloads.
- Implemented automated workflows and monitoring to improve application reliability.

SKILLS

Languages: C++, Python, C, SQL, JavaScript

Machine Learning: PyTorch, LLMs, Transformer Architectures, Diffusion Models, Deep Learning, Fine-Tuning, Model Evaluation, Generative AI

Systems Programming: Linux, UNIX, Multi-threading, Concurrency, Performance Optimization

Networking: TCP/IP, UDP, Routing Concepts, Packet Processing, Network Debugging

Distributed Systems: Scalability, Fault Tolerance, Distributed Architectures

Debugging & Profiling: gdb, Runtime Analysis, Profiling, Log Analysis, Benchmarking

Cloud & Infrastructure: AWS (EC2, S3), Docker, Kubernetes, CI/CD

Programming Concepts: Object-Oriented Programming, Data Structures, Algorithms, Software Design

EDUCATION

California State University, East Bay
Master of Science, Computer Science. *Dec 2026*

BMS College of Engineering, Bangalore
Bachelor of Engineering, Electronics & Communication. *Jun 2023*