

PKG Template

pkg@email.com | +91-9431582000 | [GitHub ↗](#) | [LinkedIn ↗](#)

PROFESSIONAL SUMMARY:

Experienced Software Engineer with 7+ years of expertise in building scalable distributed systems, cloud-native applications, and full-stack solutions. Proven track record of leading engineering teams, architecting complex systems, and delivering high-impact products at leading tech companies. Strong background in microservices architecture, DevOps practices, and agile development methodologies with focus on performance optimization and system reliability.

WORK EXPERIENCE

ANTHROPIC ↗ | STAFF AI RESEARCHER

Experience: 1.8 years | San Francisco, CA

- **Architected and deployed** large-scale inference infrastructure serving Claude API, handling **billions of tokens daily** with sub-200ms latency.
- Designed and implemented **distributed model serving system** using Ray and custom orchestration, achieving 99.99% uptime.
- Built real-time monitoring and observability platform tracking **1000+ metrics** across inference clusters using Prometheus and Grafana.
- Optimized CUDA kernels and memory management for transformer models, **reducing inference costs by 40%**.
- Developed safety evaluation pipelines for constitutional AI, processing **millions of model outputs** for alignment testing.
- Collaborated with research team to deploy experimental features, enabling **rapid prototyping and A/B testing** of new capabilities.

OPENAI ↗ | SENIOR AI RESEARCHER

Experience: 2.2 years | San Francisco, CA

- Engineered high-throughput **API gateway and rate limiting system** serving GPT models to **10M+ developers** globally.
- Built scalable embedding and fine-tuning infrastructure supporting **custom model training** for enterprise customers.
- Implemented distributed caching layer using Redis and Memcached, **reducing API latency by 60%** and improving user experience.
- Developed automated deployment pipelines for model updates, enabling **zero-downtime releases** across production clusters.
- Created comprehensive logging and tracing system for debugging LLM behaviors, processing **100TB+ logs monthly**.

GOOGLE ↗ | SOFTWARE ENGINEER II

Experience: 2.0 years | Mountain View, CA

- Developed and maintained **distributed storage systems** for Google Cloud Platform, serving petabytes of data with strong consistency.
- Built real-time data processing pipelines using Dataflow and BigQuery, enabling **analytics on 500M+ events per day**.
- Implemented advanced caching strategies and CDN optimizations, **improving content delivery speed by 70%**.
- Designed monitoring dashboards and alerting systems, reducing **mean time to detection (MTTD)** by 80%.
- Collaborated with SRE teams to improve system reliability, achieving **99.95% uptime** for critical services.

AMAZON ↗ | SOFTWARE DEVELOPMENT ENGINEER

Experience: 1.3 Years | Seattle, WA

- Designed and implemented **RESTful APIs** serving millions of requests daily for AWS management console features.
- Built distributed tracing system using X-Ray, enabling **real-time monitoring and debugging** across microservices.
- Developed automated testing framework using Pytest and Selenium, **improving test coverage by 45%**.
- Optimized **DynamoDB queries and data modeling**, reducing read/write costs by 40% while maintaining performance.

PERSONAL PROJECTS

KUBERNETES | CONTAINER ORCHESTRATION

- Contributed to open-source Kubernetes project with **multiple merged PRs** in core components.
- Implemented custom controllers and operators for automating complex deployment workflows.
- Active community member participating in SIG discussions and code reviews | [LINK ↗](#)

TENSORFLOW | MACHINE LEARNING FRAMEWORK

- Contributed bug fixes and performance improvements to TensorFlow core libraries.
- Developed custom training loops and model architectures for computer vision tasks.
- Created tutorials and documentation helping **1000+ developers** get started with ML | [LINK ↗](#)

EDUCATION

STANFORD UNIVERSITY

M.S. | Computer Science
2015 - 2017 | GPA: **3.9/4.0**

UNIVERSITY OF CALIFORNIA, BERKELEY

B.S. | Computer Science & Engineering
2011 - 2015 | GPA: **3.8/4.0**

TECHNICAL SKILLS

- Languages:** Python, Java, JavaScript, TypeScript, Go, C++, SQL
- Frameworks:** React, Node.js, Spring Boot, Django, Flask, Express.js
- DevOps:** Docker, Kubernetes, Jenkins, GitLab CI, Terraform, Ansible
- Cloud:** AWS (EC2, S3, Lambda, RDS), Google Cloud Platform, Azure
- Databases:** PostgreSQL, MySQL, MongoDB, Redis, DynamoDB, Cassandra
- Tools:** Git, JIRA, Confluence, Postman, DataDog, Grafana, Prometheus

SOFT SKILLS

- Leadership:** Team Management, Mentoring, Cross-functional Collaboration
- Agile & Project Management:** Sprint Planning, Release Planning, Retrospectives
- Process & Quality:** Code Reviews, Documentation Standards, Testing Best Practices
- Communication:** Technical Presentations, Stakeholder Management, Conflict Resolution

INTERESTS & HOBBIES

- Technical:** Distributed Systems, Cloud Architecture, DevOps, System Design, Open Source
- Community:** Tech meetups, mentoring students, contributing to open source projects
- Personal:** Hiking, photography, reading technical blogs and research papers



PyTorch

Hugging Face

spaCy

LangChain LlamaIndex

haystack

databricks

mlflow

Jina

Weaviate

vLLM PEFT
Parameter-Efficient Fine-Tuning
drant ragas