John Doe

San Francisco, CA | john.doe@email.com | github.com/johndoe | linkedin.com/in/johndoe

Professional Experience

Software Engineer, Model Infrastructure July 2024 — Present
Anthropic *San Francisco, CA*

- Designed and maintained high-throughput inference infrastructure serving Claude API traffic at scale, achieving 99.99% uptime across multi-region deployments.
- Built distributed batching and scheduling systems in Python and C++ to optimize GPU utilization across thousands of accelerators, reducing inference cost by 30%.
- Developed internal tooling for model evaluation pipelines, enabling rapid iteration across safety and capability benchmarks for pre-release model versions.
- Collaborated with research teams to productionize new model architectures, owning the full path from prototype to serving infrastructure.

Software Engineer, Market Data & Connectivity July 2023 — June 2024
Hudson River Trading *New York, NY*

- Developed ultra-low-latency market data processing pipelines in C++ handling 10M+ messages/second with sub-microsecond processing overhead.
- Optimized critical path network I/O using kernel bypass (DPDK) and lock-free data structures, reducing median tick-to-order latency by 40%.
- Built and maintained exchange connectivity adapters for equities and futures venues across US and European markets.
- Contributed to internal simulation framework used to replay historical market data for strategy backtesting and system regression testing.

Software Engineering Intern June 2022 — August 2022
Amazon Web Services *Seattle, WA*

- Built a distributed tracing feature for AWS Lambda, surfacing cold start latency breakdowns in CloudWatch, shipped to production and used by 10k+ customers.
- Wrote automated integration tests across 3 AWS regions, catching 2 critical regressions before release.

Education

Massachusetts Institute of Technology (MIT) Cambridge, MA
B.S. Computer Science & Engineering | GPA: 4.8/5.0 *September 2019 — June 2023*

Skills

Programming: C++ (expert), Python (expert), Rust (proficient), Go, Bash
Systems: Distributed systems, low-latency networking, DPDK, lock-free concurrency, CUDA
ML Infrastructure: Model serving, inference optimization, GPU scheduling, PyTorch
Tools: Git, Linux, Docker, Kubernetes, AWS (EC2, Lambda, CloudWatch)
Languages: English (Native), Spanish (Conversational)