Phong Yang

contact@phongyang.com (320) 201-4824 https://phongyang.com

SUMMARY

Senior Software Engineer with 15 years of experience building scalable backend systems, distributed platforms, and cloud-based applications. Experienced in Java, Python, Go, and TypeScript with a strong background in REST APIs, microservices, search infrastructure, and AWS cloud architecture. Recently focused on AI-powered and media processing platforms, integrating machine learning inference, video processing pipelines, and high-performance distributed systems. Proven track record of improving system scalability, performance, and reliability across enterprise and startup environments.

EDUCATION

University of Minnesota, *Master of Science in Computer Science* 2014 – 2016

University of Minnesota, Twin Cities, 2006 – 2010
Bachelor of Science in Computer Science

WORK EXPERIENCE

Twelve Labs, *Senior Software Engineer* 09/2022 – Present

Technologies: Go, Node.js, TypeScript, React/Next.js, GraphQL, AWS (ECS/Fargate, Lambda, Step Functions, Media Services), Postgres, Redis, Kafka/SQS, Docker/Kubernetes, Terraform, FFmpeg/HLS

- Designed and implemented scalable RESTful and GraphQL APIs adhering to OpenAPI specifications, powering features such as video search, embedding generation, and multimodal model inference, enabling low-latency access to Twelve Labs' video understanding platform.
- Architected high-throughput, service-oriented backend systems in Node.js and Go, deploying via AWS ECS/Fargate and Kubernetes, to support enterprise-grade SaaS workloads and multi-tenant video analytics pipelines.
- Optimized distributed video data processing with FFmpeg-based transcoding and AWS Media Services, achieving sub-second inference latency and improving system throughput by 45% through asynchronous, event-driven workflows (SQS, Kafka).
- Built responsive frontend interfaces in Next.js and React (TypeScript, react-query, Tailwind CSS), providing real-time visualization of video embeddings, search results, and annotation timelines connected to backend inference APIs.
- Integrated AI/ML pipelines into user-facing applications, exposing model outputs (captions, embeddings, temporal highlights) through REST and streaming endpoints, and collaborating closely with Al teams to ensure data consistency and scalability.
- Developed resilient background processing systems using BullMQ, Redis Streams, and Step Functions to orchestrate long-running video operations such as encoding, similarity computation, and indexing.
- Implemented observability and monitoring stack with Prometheus, Grafana, Datadog, and OpenTelemetry, establishing SLOs and automated alerting that reduced mean time to recovery (MTTR) by >30%.
- Led modernization of legacy prototype systems into modular, containerized microservices with automated deployment pipelines using GitHub Actions, Terraform, and AWS CDK, supporting blue-green deployments and A/B model testing.

- Collaborated cross-functionally with designers and PMs using Figma, ensuring seamless UX integration with backend APIs; improved first meaningful paint by ~35% through code-splitting and caching optimization.
- Mentored engineers on API design, distributed architecture, and modern frontend best practices, helping establish shared component libraries and documentation aligned with Twelve Labs' product evolution.

Intuit, *Senior Software Engineer* 06/2018 – 09/2022

- Implemented backend features and incremental enhancements across VEP growth/retention services using Java/Python, improving reliability and consistency of expert and customer workflows.
- Built and maintained RESTful APIs (and/or GraphQL endpoints where applicable), ensuring clear contracts, input validation, and backward compatibility for downstream clients.
- Diagnosed and resolved defects across QA, pre-production, and production, delivering timely hotfixes and post-release patches.
- Wrote unit tests and applied test-driven development (TDD) to reduce regressions and improve deployment confidence.
- Participated in architecture discussions and contributed to scalable microservice design.
- Collaborated in Agile/Scrum environment with engineers, product managers, and designers to deliver iterative improvements.
- Contributed to AI-driven experiences by integrating backend services with NLP / ML / generative AI capabilities.

Intuit, *Software Engineer II* 06/2016 – 06/2018

- Implemented backend services and APIs using Java and Python for customer and expert platform workflows.
- Built RESTful services with strong input validation, logging, and monitoring.
- Fixed production issues and implemented performance improvements.
- Wrote unit tests and participated in code reviews and engineering documentation.
- Worked in Agile environment delivering incremental platform improvements.

Thomson Reuters, *Software Engineer* 06/2010 – 08/2016

Technologies: Java, Spring, Python, REST, SOAP, Oracle, SQL Server, AWS (EC2, S3), Jenkins, AngularJS, Solr, Elasticsearch, Agile/Scrum

- Developed and maintained backend services in **Java (Spring)** and **Python** supporting Thomson Reuters legal and tax research platforms serving enterprise customers.
- Designed and implemented **RESTful APIs** and maintained legacy **SOAP services**, enabling integration with internal and external enterprise applications.
- Built and optimized **SQL queries and database schemas (Oracle, SQL Server)**, improving large-scale document retrieval performance by approximately **30%**.
- Implemented **search infrastructure using Apache Solr**, improving search relevance and reducing query response times for legal document search systems.
- Contributed to early **cloud migration initiatives to AWS (EC2, S3)**, helping transition legacy on-premise applications to more scalable infrastructure.
- Developed internal web tools using **AngularJS**, enabling editorial and data teams to manage and review legal content more efficiently.
- Set up **Jenkins CI pipelines** and automated testing workflows, reducing deployment issues and improving release reliability.
- Worked in an **Agile/Scrum** environment, participating in sprint planning, code reviews, and cross-team architecture discussions.
- Collaborated with product managers and data teams to build **large-scale content processing and indexing pipelines** for legal and financial documents.