Hoang Nguyen

+1 361-419-4525 | hoangndst@gmail.com | linkedin.com/in/hoangndst | github.com/hoangndst | hoangndst.com

Technical Skills

Languages: Go, Python, C++, TypeScript, JavaScript, SQL, Java, Bash, Dart
AI/ML: PyTorch, TensorFlow, RAG, LLM Evaluation, MLOps, Google ADK, MCP
Systems/Delivery: Linux, Kubernetes, Cluster API, Docker, Terraform, Helm, Kustomize, Jenkins, Argo CD, CI/CD, GitOps
Observability: OpenTelemetry, Prometheus, Grafana
Cloud: AWS, GCP, Azure, Cloudflare
Data: PostgreSQL, Redis Streams, pgvector
Frameworks/Testing: FastAPI, React, Next.js, Flutter, REST APIs, gRPC, pytest, Playwright
AI Coding Tools: Codex, Claude Code, Cursor, Antigravity

Experience

Graduate Research Assistant September 2025 – Present
The University of Texas at San Antonio *San Antonio, TX*

- Built an LLM evaluation harness with deterministic holdout splits, **token F1**, **retrieval hit@k**, **p50/p95** latency, token usage, and provider-error reporting
- Enabled project-scoped AI assistance by implementing **hybrid RAG** with **PostgreSQL** full-text search, **pgvector** similarity search, reciprocal-rank fusion, and streamed responses
- Automated personalized messaging for approximately **20,000 users** by building **Python**/FastAPI APIs and graph-based runtime with conditional branching and participant-specific scheduling
- Resolved delayed, mistimed, and duplicate delivery by implementing a **PostgreSQL** transactional outbox and idempotent **Python** workers on **Redis Streams**, enabling horizontal scaling
- Enabled researchers to validate **multi-day** messaging protocols in **minutes instead of days** before live delivery by building a **TypeScript**/React console with **virtual-clock simulation** of the full protocol program

Software Engineer January 2023 – May 2025
Viettel Group *Hanoi, Vietnam*

- Cut mean time to recovery by **40% across 1,000+ clusters** by architecting a centralized observability stack for metrics, logs, traces, alerts, and application SLIs
- Delivered **99.9% uptime** by launching production **Cluster API** components, custom **Go** controllers, and a Secrets Store CSI provider with KMS-backed credential mounts
- Cut **Kubernetes** upgrade time by **15%** by building a **Cluster API** operator for rolling/blue-green node replacement, workload draining, and rollback; presented the design at FOSSASIA Summit
- Cut multi-environment provisioning time by **40%** with **Terraform**, **Kustomize**, and **Helm**; enabled **5 enterprise clients** to self-serve infrastructure and application delivery through the Kusion platform
- Increased deployment frequency by **25%** by migrating SInvoice to **Kubernetes**, automating **Argo CD** deployments, and building queue-driven, autoscaling Jenkins CI/CD runners

Software Engineer Intern August 2022 – December 2022
Viettel Group *Hanoi, Vietnam*

- Cut real-time video analytics latency by **10%** at **50,000+ concurrent feed connections** by optimizing **Python**/**Go** services and data pipelines
- Increased weekly active users by **12%** and monthly retention by **8%** within three months by building a **Next.js** video surveillance platform on load-balanced infrastructure

Projects and Leadership

Spektro | *Independent Project; Go, Google ADK, MCP* getspektro.com

- Built an incident-RCA platform that correlates monitoring, logs, inventory, and tickets into a root cause in seconds and drafts MOP/EOP/SOP documentation by designing **8 Go/TypeScript microservices** with a **gRPC** AI investigation pipeline for evidence admission, trust scoring, and model routing
- Shipped **3** Go services as hardened **non-root/distroless** container images; secured funding and signed the **first enterprise customer**

Kusion (Maintainer) | *CNCF Sandbox; Go, Terraform, Kubernetes* github.com/KusionStack/kusion

- Maintain **Go** orchestration components in the **1.3k-star** project; co-presented Kusion IDP patterns with project contributors

ezmsg (Research) | *Messaging Protocol System; Python, Next.js, Flutter* dashmessaging.hoangndst.com

- Built an end-to-end messaging-protocol management system serving approximately **20,000 users** by shipping a **Python**/FastAPI backend, **Next.js** operator console, and **Flutter** mobile application

Conference Speaker | *5 editions: Conf42 2024, FOSSASIA 2024/2025, OpenInfra NA 2024, SREday Austin 2026*

Education

The University of Texas at San Antonio San Antonio, TX
Master of Science in Computer Engineering (GPA: 3.91) *September 2025 – May 2027*

- Expected graduation: May 2027; fully funded competitive scholarship

Vietnam National University Hanoi, Vietnam
Bachelor of Engineering in Computer Engineering *October 2020 – December 2024*