Le Xuan Cuong
cuongkane.com

Email: cuongle.software@gmail.com
Mobile: +84 77 888 7894

SUMMARY

Senior Backend Engineer with 6 years of experience building highly available, scalable cloud systems processing millions of daily transactions. A collaborative builder and critical thinker who values simplicity and low-risk solutions. Led architecture design for 10+ technical designs across microservices, Machine Learning systems, and real-time pipelines. Combines an innovative, highly adaptive mindset with deep engineering accountability. Active technical blogger at cuongkane.com.

EXPERIENCE

Parcel Perform — Ho Chi Minh City, Vietnam
Senior Backend Engineer — *Aug 2024 - Present*

- **System Architecture & Design**: Directed architecture for 10+ systems including ML and real-time pipelines; authored cross-team Technical Design for carrier identification spanning 7 microservices, approved by CTO.
- **Microservices & Database Splitting**: Proposed and Engineered decomposition of 2 monoliths into microservices with zero-downtime database splitting, enhancing modularity and enabling autonomous team deployments, improving team velocity by 30%.
- **High-Availability Cloud Systems**: Proactively take over and maintain mission-critical analytics platform (ClickHouse/Dagster); accelerated dashboard chart loading by implementing static file caching strategies.
- **Real-Time Data Pipelines**: Optimized ML-based carrier identification pipeline processing millions of daily events across 7 microservices; increased success rate by seamlessly integrating ML models with legacy regex flows.
- **AI/ML Solutions**: Pioneered building AI agents using LangChain following ReAct architecture and AWS Bedrock; reduced LLM costs by 20% via strategic model selection while automating Root Cause Analysis workflows.
- **Engineering Excellence**: Pioneered company-wide adoption of AI-assisted Technical Design workflows, ensuring code quality with LLM.
- **Release & Deployment Strategy**: Implemented advanced release strategies including canary deployment protocols and A/B testing, achieving low-risk delivery through incremental design principles.
- **Observability & Sustainable Growth**: Enhanced system reliability by defining and monitoring SLA/SLO targets; utilized percentile latencies and Kafka lag metrics to drive proactive capacity planning and ensure consistent service availability.
- **Mentorship & Knowledge Sharing**: Accelerated team growth by onboarding 5+ engineers and mentoring a developer to promotion-readiness; fostered engineering excellence by delivering 15+ company-wide technical sessions on System Design, AI/LangChain, and best practices.
- **Engineering Leadership**: Strengthened engineering standards by sharing culture, showcasing clean code principles and software best practices, and conducting 6+ technical interviews.

Parcel Perform — Ho Chi Minh City, Vietnam
Backend Engineer — *Jul 2022 - Aug 2024*

- **Engineering Excellence**: Maintained stable and fast development velocity (compared to other devs) by applying best practices (TDD) and effective task management.
- **Scalable API Development**: Designed and Launched Public Booking API for multi-carrier integrations; and Celery-based rate limiting to protect third-party resources.
- **Extensible Integration Architecture**: Designed extensible integration framework supporting 50+ carriers; validated requirements through usage testing to ensure scalability for future logistics partners.
- **High-Volume Data Pipelines**: Scaled data pipeline to process 5M+ daily events for mission-critical tracking using Apache Flink and optimized microservices.
- **Email Extraction Platform**: Engineered high-throughput email extraction engine using Kafka CDC; system has processed shipment notifications reliably in production for 3+ years.
- **Sync-to-Async Reliability Pattern**: Ensured booking reliability by architecting a sync-to-async fallback pattern across 5 microservices using Kafka to handle component failures.
- **Celery Pipeline Architecture**: Accelerated data processing by parallelizing 8K+ postal code tasks using optimized Celery canvas patterns (chain/chord) and worker configurations.

- ○ **System Reliability & Incident Response**: Restored 1.5M records during critical incidents within 24 hours by developing custom recovery scripts and resolving Redis cluster hotspots under 10x traffic load.
  - ○ **Large-Scale Data Migrations**: Secured data consistency during large-scale migrations by implementing incremental update patterns, optimizing SQL indexing, and managing autovacuum constraints.
  - ○ **Testing & Developer Experience**: Defined and spread company-wide testing standards (AAA pattern) and drove adoption of modern tooling (Poetry, Ruff).
- **Cinnamon AI Labs**   Ho Chi Minh City, Vietnam
  Software Engineer   *Aug 2020 - Jul 2022*
  - ○ **Enterprise Client Delivery**: Deployed enterprise-grade document processing systems for large enterprise clients (Deloitte, Bridgestone,...), directly driving client success in AI automation.
  - ○ **Multi-Framework Backend Development**: Developed versatile backend solutions using Flask, FastAPI, and Django, strictly adhering to SOLID principles and Repository patterns.
  - ○ **Multi-Cloud Infrastructure**: Managed multi-cloud infrastructure (AWS, Azure, GCP) and containerized applications with Docker to ensure consistent cross-cloud deployments.
  - ○ **ML-to-Production Pipeline**: Bridged gap between research and production by converting experimental ML models into scalable, SLA-compliant API services.

PROJECTS

- **Codlis - AI Code Visualization (Master's Thesis, H2 2025) (codlis.com)**: Architected multi-agent LLM system for interactive code comprehension using LangChain and Knowledge Graphs. Validated effectiveness via user study showing 2.9x higher task success rate and 29% cognitive load reduction.
- **SweatCharge - Sports Club Management (H1 2025) (sweatcharge.com)**: Developed full-stack sports management platform with role-based SSO and automated billing; currently servicing active production users. Implemented fuzzy matching algorithms for player association using React/TypeScript.
- **Hasiti - AI Hairstyle Simulator (Bachelor's Thesis, 2021)**: Built real-time hairstyle simulation app using fine-tuned CNNs for style transfer. Architected full-stack solution connecting React Native mobile client with Python/Flask AI inference backend.

CERTIFICATIONS

- **AWS Certified Solutions Architect - Associate**: Amazon Web Services (Valid: Jan 2025 – Jan 2028)

EDUCATION

- **Master of Science in Computer Science**   Sept. 2021 – Sept. 2025
  HCM University of Science   *Ho Chi Minh City, Vietnam*
  - ○ **Research Focus**: AI/ML, Big Data, and Human-Computer Interaction under post-doctoral supervision. Applied analytical experimentation methods for metric-driven evaluation.

- **Honor Program - Bachelor of Science in Information Technology**   Sept. 2017 – Sept. 2021
  HCM University of Science   *Ho Chi Minh City, Vietnam*
  - ○ **Honor Program**: Top 70 students selected from 1,000+ candidates. Comprehensive curriculum spanning web/mobile development, computer vision, image processing, software engineering, and scientific research methods.

  - ○ **Technical Foundation**: Developed fundamentals in algorithms, data structures, OOP principles, and hardware/software architecture through rigorous coursework.

- **Machine Learning Course (Facebook-sponsored)**   Aug. 2019 – Dec. 2019
  Coder School   *Ho Chi Minh City, Vietnam*

AWARDS & RECOGNITION

- **Parcel Perform (2022-Present)**: Member of the Month (3x), Best Presentation Award, AVA AI Award for AI adoption, Professional Style Award.
- **Cinnamon AI Labs (2020-2022)**: Employee Recognition Award (Monthly) for outstanding technical contributions.
- **Faculty of IT Academic Excellence (2018)**: Top 10 students with highest GPA (9.08/10) among 1,000+ students.

Technical Skills

- **Architecture**: High-Availability Systems, Microservices, System Design, RESTful APIs, TDD, Idempotency, Outbox Pattern, Feature Toggles, SSO/OAuth

- **Big Data & Pipeline**: Apache Flink, Kafka, Celery (canvas: chain/chord/group), Dagster, RabbitMQ

- **Infrastructure & Observability**: AWS (EC2, S3, Bedrock), Docker, Grafana, Sentry, Alerting Systems, CI/CD

- **AI/ML**: LangChain, LangGraph, LangSmith, AI Agents, Prompt Engineering

- **Languages**: Python, Scala, Java, TypeScript, SQL

- **Databases & Data Stores**: PostgreSQL, Redis, ClickHouse

- **Web Frameworks**: Django, Django Ninja, FastAPI, Pydantic, Next.js/React (side projects)

- **Engineering Practices**: Technical Planning, Task Management, Story Estimation, Sprint Planning, Agile

- **Dev Environment**: Vim (primary editor), tmux, terminal-centric workflow, split keyboard enthusiast

Languages

- **English**: B2 Upper Intermediate – Aptis ESOL (British Council) 176/200   **Vietnamese**: Native