

# AUGUSTINE NGUYEN - CURRICULUM VITAE

---

Data Scientist · Financial Risk & Growth Analytics · Computer Vision

Republic of Korea · [augustino0890@gmail.com](mailto:augustino0890@gmail.com) · [LinkedIn](#) · [GitHub](#) · [Twitter](#)

| Where credit intelligence meets customer growth — turning data into revenue and trust.

---

## Profile

Data Scientist with extensive industry experience across **credit risk modeling, fraud detection, growth analytics, generative AI, and computer vision**. Proven record of shipping production ML systems serving 10M+ users, architecting data platforms handling 5B+ records, and winning a global hackathon championship in enterprise RAG. Fluent in both the quantitative rigor of financial risk and the experimental culture of growth analytics, with a growing focus on perception systems where modeling meets the physical world.

---

## Awards & Recognition

### Champion — FPT AI Hackathon 3 · 2025

Global competition across all FPT offices worldwide. Team *Cảm Y Vê (FKR-KTX)* placed 1st in the *KT/KB Technical Solutions* track with **AISOL**, a Japanese-language enterprise RAG platform built on Qdrant, Gemini, and Qwen.

---

## Professional Experience

### FPT Software Korea — *Data Software Engineer*

June 2024 – Present · South Korea

#### QA/QC Data Analytics and Visualization Dashboard

Developed a data analytics and visualization dashboard to monitor and analyze QA/QC data, enabling real-time insights for product quality improvement through statistical processing and visual reporting.

- Designed and implemented data pipelines to aggregate, preprocess, and analyze large QA/QC datasets using SQL and Java (Apache Commons Math), delivering statistical insights for quality enhancement.
- Applied advanced statistical methods to identify trends and anomalies, creating interactive visualizations (pie, bar, line charts) with Chart.js for stakeholder decision-making.
- Developed optimized REST APIs using Spring Boot to deliver real-time statistical reports and dashboards, streamlining data-driven quality process improvements.

*Technologies: SQL · Java · Oracle · Apache Commons Math · Chart.js · REST APIs · Spring Boot*

#### Investment Securities Project — Meta

Migrated an Oracle database to EDB (PostgreSQL), optimizing SQL queries for performance and compatibility to meet business requirements.

- Converted and optimized Oracle SQL queries to ANSI-standard for EDB (PostgreSQL) compatibility, improving query efficiency.
- Translated complex business requirements into high-performance SQL queries, ensuring alignment with project goals.

*Technologies: Oracle · EDB (PostgreSQL) · ANSI SQL*

---

### SupperTree — *Software Engineer*

April 2021 – June 2024 · South Korea

#### BlockStream — Real-Time Blockchain Data System

- Built scalable real-time monitoring for Ethereum and Polygon transaction data using a publish-subscribe architecture.
- Engineered data pipelines with Kafka, persisted in PostgreSQL, indexed in OpenSearch/Algolia for high-performance search.

- Managed end-to-end workflow from on-chain collection to storage and indexing, including enriched metadata.

### Discord Emotion Tracker

- Built an AI-powered application analyzing Discord customer service communications with real-time sentiment analysis.
- Applied NLP via AWS Comprehend and HuggingFace Rust-BERT to detect emotion patterns at scale.
- Implemented real-time English-to-Korean translation using AWS Translate for cross-market insights.

### NFT Management — Token Staking and Earning App

- Developed scalable backend infrastructure for NFT staking and token earning, ensuring reliable UX and transaction integrity.
- Integrated blockchain technology for transparent NFT transactions; optimized database queries for high-performance data handling.

*Technologies: Python · Rust · Golang · TypeScript · Node.js · Kafka · AWS Comprehend/Translate · HuggingFace · Algolia · OpenSearch · PostgreSQL · MongoDB · Kinesis*

---

### FitPetMall — *Data Software Engineer*

*July 2020 – April 2021 · South Korea*

### InsightFlow — Customer Behavior Analytics Platform

- Developed and optimized data models and workflows for logging and analyzing customer behavior on e-commerce.
- Built high-performance data processing with an analytics engine and real-time streaming via Kafka and ELK Stack.
- Architected data-driven solutions for shipping, product delivery, and promotional analytics.

*Technologies: Python · FastAPI · ELK Stack · Apache Kafka · Oracle*

---

### AZEN Global — *Data Scientist / ML Engineer*

*March 2019 – July 2020 · South Korea*

### ABACUS — Advanced Credit Risk Prediction Analysis Tool

A machine-learning platform that augments fault detection and refines credit risk modeling, sharpening the precision of financial assessments and operational reliability for an installed base of more than 10 million users.

- Developed advanced credit risk models using machine learning algorithms, achieving a 3% lift in scoring accuracy and improving the precision of downstream financial assessments.
- Optimized fault-detection data pipelines by streamlining processing on Apache Spark, materially increasing throughput for workflows serving 10M+ users.
- Architected highly-available, resilient database clusters across ClickHouse, MongoDB, and Redis to manage 5B+ records, with disaster-recovery provisions safeguarding data integrity and operational continuity.
- Built real-time monitoring dashboards and interactive reports surfacing live platform signals, strengthening operational transparency and supporting informed decision-making.

*Technologies: Python · Django · FastAPI · ClickHouse · MongoDB · Oracle · Redis · Apache Spark*

---

### National Cancer Center — *Graduate Research Assistant (Health Data Analyst)*

*September 2016 – September 2018 · Goyang, South Korea*

### Mendelian Randomization — Causal Analysis of BMI on Thyroid Cancer Risk

- Conducted statistical analyses to assess causal relationships between BMI and thyroid cancer risk using Python, R, and SAS.
- Integrated genome-wide SNP data with environmental factors to identify causal genetic relationships.
- Applied instrumental variable analysis (Mendelian Randomization) — methodology directly transferable to financial econometrics and policy evaluation.

*Technologies: SAS · Python · R · Linux*

---

## Education

### Master of Public Health — Health Data Science

National Cancer Center, Goyang, Republic of Korea · 2016 – 2018

Thesis: *Causal Effect of BMI on Thyroid Cancer Risk using Mendelian Randomization*

### Bachelor of Engineering — Biotechnology

Industrial University of Ho Chi Minh City, Vietnam · 2008 – 2013

---

## Credentials

- AWS Certified Data Engineer — Associate
  - AWS Certified AI Practitioner — Foundational
  - Microsoft Certified: Azure AI Fundamentals
  - Nanodegree: Data Engineering with Microsoft Azure
  - Astronomer Certification: Apache Airflow Fundamentals
- 

## Technical Skills

DOMAIN	METHODS & TOOLS
Generative AI & RAG	LLM Pipelines · RAG · Vector Search · Qdrant · Gemini · Qwen · LangChain · Prompt Engineering
Computer Vision	Object Detection (YOLO · SSD) · Multi-Object Tracking (DeepSORT · ByteTrack · StrongSORT) · Re-Identification · TensorRT · OpenCV · Edge Inference (Jetson) · ROS 2
ML & Statistical Modeling	Scikit-learn · XGBoost · LightGBM · Statsmodels · Survival Analysis (Kaplan-Meier · Cox PH · AFT) · SHAP · LIME
Financial Analytics	Credit Scoring · Fraud Detection · Risk Dashboards · Basel Compliance Metrics
Growth & Marketing	RFM · Cohort Analysis · CLV · MMM · Funnel Analytics · Propensity Scoring
Causal & Experimental	A/B Testing · Bayesian A/B · Thompson Sampling · IV · DiD · PSM · CUPED
Data Engineering	SQL · Pandas · dbt · Apache Spark · Kafka · Airflow · ELK Stack
Databases	PostgreSQL · Oracle · ClickHouse · MongoDB · Redis · OpenSearch
Cloud & Infrastructure	AWS · Azure · Docker · Kubernetes
Languages	Python · Rust · SQL · R · Golang · SAS · TypeScript · Java

---

## Open Source & Side Projects

- **discord-playdapp-bot** — Rust · Serenity · MongoDB · Docker

High-performance Discord bot built in Rust using Serenity for gaming community management. Handles tournament ticket exchange, player engagement tracking, and leaderboard ranking via slash commands.

- **outfit-square** — Python · Discord.py · MongoDB · Poetry

Python-based Discord bot for user point management in metaverse gaming channels. Integrates MongoDB for persistent storage with multi-stage (dev/prod) deployment and OAuth2 authentication.

---

## Personal Details

- Citizenship: Vietnamese
- Country of Residence: Republic of Korea
- Languages: English (fluent) · Vietnamese (native)