

Long Vu Van

+84 328 828 140 | vulong3896@gmail.com | [in LinkedIn](#) | [📍 Ha Noi, Viet Nam](#)

EDUCATION

University of Engineering and Technology - Vietnam National University <i>Master degree in Computer science - working on master thesis</i>	Hanoi, Vietnam <i>Apr. 2023 – 2025</i>
Hanoi University of Science and Technology <i>Engineering degree in Information and Technology - Talent program</i>	Hanoi, Vietnam <i>Jun. 2014 – May 2019</i>

EXPERIENCE

ML Product Manager <i>CMC ATI - an AI applied technology institute</i>	Feb. 2023 – Present <i>Hanoi, Vietnam</i>
<ul style="list-style-type: none">• Roadmap alignment: Develop year/quarter/month plan that aligned with company mission and product strategy. Prioritized initiatives based on business impact, resource availability, and customer's business need.• Task delegation & Tracking: Assigned tasks to team members using SMART goals (Specific, Measurable, Achievable, Relevant, Time-bound). Matched responsibilities to individual strengths and project requirements.• Performance review: Monitored deliverables via Jira/GitLab MRs (code quality, timeline adherence) and verbal syncs (to solve blockers, collaboration issues). Conducted retrospectives to refine processes and address gaps in execution.• Research strategy: Market research from firms like Gartner along with business requirements from both internal and external to bring out the detailed machine learning research plan• DevOps & Infrastructure: built an on-premise K8s cluster for deploying AI models and ML software products. Implemented the following technologies and workflows on the cluster: CI/CD pipeline, secret management system, database operator, proxy server for routing request, service monitoring• MLOps: Implemented an MLOps system level 3 including model registry, data drift detection, automated training, data version control (DVC) and CI/CD model pipeline• Customer centric: Worked closely with customer to bring business ideas into machine learning models and software features and do project management to deliver the product to them. Consulted the customer to fulfill the gap between business expectation/requirement with the AI technology• Human resource management: provide clear career path for each position in the team such as techstacks/softskills they need to learn. Led end-to-end recruitment for technical roles: screened resumes, conducted technical interviews, and evaluated cultural fit. Focused on scalable hiring—balancing immediate project needs with long-term team growth.	
Freelancer MLOPs Engineer <i>Alethea - an NFT startup</i>	Jan. 2022 - Aug. 2023 <i>United States</i>
<ul style="list-style-type: none">• Deployed a voice service with these components: text to speech/speech to text service, Redis memcache (self-hosted), Triton inference server on AWS EKS, AWS ECR and a VPC peering to synchronize the result to many other VPCs in the fastest way.• Deployed an automated pipeline for voice training and model CI/CD using Kubeflow, K8s, AWS EKS, and Triton inference server which reduces the time to deploy a new voice by double and also reduces the cloud cost.• Wrote data pipelines to process text to speech data using NeMo - a tool of NVIDIA to process speech data, this pipeline is hosted in Kubeflow.	
Solution Architect <i>akaBot FPT Software - an RPA startup</i>	Apr. 2021 - Dec. 2022 <i>Ha Noi, Viet Nam</i>
<ul style="list-style-type: none">• Deployed an AWS EKS cluster for serving models using ModelMesh Serving framework which is more flexible, easy to horizontally scale• Designed a distributed task system which is asynchronous, elastic and scalable by combining Celery, a message queue to send and receive result using RabbitMQ, a task dispatcher using FastAPI, and a data storage system using AWS S3. Deployed it to cloud to serve customer on AWS EKS.• Developed an information extraction service that's being used on Europe financial reports. Deployed it to cloud to serve customer on AWS EKS.• Optimized the AI models by using model pruning, model quantization, TensorRT library, ONNX format and Triton model serving.	

- Implemented a Robotic Process Automation selector optimization algorithm which minimizes selector nodes and attributes by using tree traversal algorithm and heuristic algorithm.
- Developed a computer vision set of activity which automates human task by analyzing computer screen information using object detection and optical character recognition.

AI Engineer

May. 2020 - Apr. 2021

Asilla Vietnam - an AI startup specializing in image and video analysis

Ha Noi, Viet Nam

- Built and deployed overall video streaming and broadcasting system which reads camera input stream, then process it by Computer Vision plugin and then broadcast the output stream to many registered devices based on a media streaming server named Kurento and Spring Boot framework.
- Implemented a human counting program which reads the camera's streams and do many actions like human counting, trespassing detection based on object detection. The final test achieved over 90 percent of accuracy on the real scenarios.
- Implemented an age gender recognition program which achieved 96 percent of gender accuracy and 2 Mean Average Error on age recognition, overcome the AWS's age recognition service which achieved 4 MAE on a same test.

AI Engineer

Feb. 2019 - May. 2020

CMC CIST - an AI research institution

Ha Noi, Viet Nam

- Implemented a smart transportation supervision system including vehicle detection, license plate detection, license plate recognition, and violation detection.
- Researched human facial recognition on the Vietnamese facial dataset, the final result achieved 99.8 percent of accuracy on 1000-samples dataset which is the result of many improvement techniques like data augmentation, model structure modifications, and data engineering.

PROJECTS

Smartdoc - an IDP application | *Python, Django, Celery, Angular, PostgreSQL*

Feb 2023 – Present

- Designed and developed a full-stack web application using with Django with Angular as the frontend
- Used by thousands daily, the product has processed millions of documents for Vietnamese local governments in many digitalization projects
- Used Celery, Redis and RabbitMQ for asynchronous tasks; Used Flower for task monitoring
- Used Kubeflow to automate model training and model CI/CD pipeline
- Used OCR, document classification, document instance recognition, information extraction, LLM in the AI pipeline

TECHNICAL SKILLS

Languages: Python, JavaScript, TypeScript, SQL

Technologies: Flask, FastAPI, Django, Celery, React.js, AngularJS, Git, Kafka, RabbitMQ, Redis, etcd, OpenCV, PyTorch, Infisical

DevOps: Docker, AWS, Kubernetes, Ansible, Kubespray, DVC

MLOps: MLflow, Kubeflow, Triton, Model mesh serving, model registry, feature store

Database: MySQL, PostgreSQL, MongoDB, Elasticsearch, Milvus, Neo4J

Methodologies: Agile, Scrum, OOP, Functional Programming, DevOps, CI/CD, TDD, SOLID

Others: Jaeger request tracing, Grafana, Prometheus

CERTIFICATES

Language: TOEIC 905 (2022), VSTEP 7.5 (2023)

Cloud: AWS SAA C03 (2023)