Quan Minh Bui

AI Scientist | Cloud Engineer | NLP Researcher
Address: 336-0911 埼玉県, さいたま市緑区大字三室1245-5
Phone: +81 80-8854-2577 | Email: quanbm2710@gmail.com

Professional Summary

AI Scientist and Cloud Engineer with a strong background in Natural Language Processing, Large Language Models (LLMs), and AWS-based AI system design. Experienced in research and industry projects, including LLM distillation, weak supervision, RAG systems, and AI-powered automation. Skilled in developing scalable architectures, deploying ML models on AWS, and publishing academic papers in top AI conferences and journals.

Core Skills

• Programming: Python | Flask | FastAPI
• AI/ML Frameworks: TensorFlow, PyTorch, Lightning AI
• NLP Tools: HuggingFace, Spacy, AllenNLP, LangChain
• Cloud & DevOps: AWS (API Gateway, Lambda, SageMaker, Step Functions, CodePipeline, EC2, VPC, etc.)
• LLM APIs: OpenAI, Anthropic Claude, Gemini
• Data Science: NumPy, Pandas, Matplotlib, Gradio, Streamlit
• Containerization: Docker | AWS ECS | ECR | Fargate
• IaC: AWS CDK | Terraform | TF-CDK
• Research: Experiment Design | Academic Writing | NLP Evaluation
• CV & Audio: OpenCV | Whisper | SpeechBrain | Pyannotate

Professional Experience

AI Consulting and Cloud Engineer | SBI Life Insurance Company
Jan 2024 – Present
Leading AI and cloud architecture initiatives, developing scalable AI-powered solutions using AWS services.

Senior AI Engineer | Amifiable Inc
Nov 2023 – Present
Designing LLM-powered tools and automation systems for enterprise clients, focusing on UI testing and knowledge-driven applications.

Research Assistant | JAIST - Research Center for Interpretable AI
Sep 2020 – Oct 2023
Conducted research in NLP, legal text processing, and explainable AI. Contributed to COLIEE competitions and academic publications.

Research Assistant | RICOH Joint Project
Sep 2021 – Sep 2023
Collaborated with RICOH on interpretable AI research, focusing on weak supervision and LLM distillation techniques.

Selected Projects

RAG System
Designed and implemented a Retrieval-Augmented Generation (RAG) system using AWS Kendra, Aurora, OpenSearch, and Bedrock for enterprise knowledge bases.

Private Microservice Architecture
Built a secure private microservice using AWS API Gateway, Route 53, Lambda, and Resource Access Manager (RAM).

Teams Meeting Summarization
Developed an AI-powered summarization system using Whisper for transcription and Bedrock for context-aware summarization.

Smart Locker System
Integrated AWS SES, SNS, Lambda, and Step Functions with IoT smart padlocks for secure delivery automation.

LLM Distillation
Researched and developed model distillation pipelines for domain adaptation with reduced computational cost.

Weak Supervision
Designed labeling functions and used LLMs to enhance weak supervision pipelines for NLP data annotation.

Machine Reading Comprehension with Knowledge Graphs
Constructed knowledge graphs and implemented efficient query mechanisms for question answering tasks.

Education

Ph.D. in Information Science
Japan Advanced Institute of Science and Technology (JAIST) | Sep 2023

Master in Information Science
Japan Advanced Institute of Science and Technology (JAIST) | Sep 2020

Bachelor in Computer Science
University of Information Technology | Oct 2016

Publications

All original paper titles retained for accuracy.

1. JNLP @COLIEE-2023: Data Argumentation and Large Language Model for Legal Case Retrieval and Entailment (2023)
2. JNLP Team: Deep Learning Approaches for Tackling Long and Ambiguous Legal Documents in COLIEE 2022 (Springer, 2022)
3. SM-BERT-CR: a deep learning approach for case law retrieval with supporting model (AI & Law, 2022)
4. How Curriculum Learning Performs on AMR Parsing (IEEE, 2021)
5. ALQAC 2022: A Summary of the Competition (IEEE, 2022)
6. A Summary of the ALQAC 2021 Competition (IEEE, 2021)
7. Transformer-Based Approaches for Legal Text Processing: JNLP Team-COLIEE 2021 (Review of Socionetwork Strategies, 2022)
8. Evaluate and Visualize Legal Embeddings for Explanation Purpose (IEEE, 2021)
9. How state-of-the-art models can deal with long-form question answering (PACLIC, 2020)

Awards & Scholarships

- First Place – COLIEE 2020, 2021, 2023 (Best system in Legal Case Retrieval and Statute QA)
- Runner-up Student Paper Award – KSE 2021
- Best Presentation Award – KSE 2021
- Excellence Award – JSAI National Conference 2025
- Doctoral Research Fellow (DRF) Scholarship – JAIST

Certifications

- AWS Certified Machine Learning – Specialty