Hieu Nguyen

Hanoi, Vietnam • hieunv.dev@gmail.com • (+84) 372615855 • linkedin.com/in/behitek

github.com/behitek • behitek.com

WORK EXPERIENCE

AI Engineer   **Feb 2023 - Present**
VMO Holdings   **Hanoi, Vietnam · On-site**
- Delivered multiple AI projects to production, including a recommendation system, RAG chatbot, and information retrieval systems, with two projects recognized as 'Project of the Year'.
- Evaluated the feasibility of customer problems, researched potential solutions, designed architecture, and anticipated potential issues, ensuring smooth project execution and delivery.
- Enhanced the accuracy of the document retrieval system by 30% compared to the baseline version.
- Drove project success by proposing innovative solutions, mentoring junior engineers, and ensuring the timely completion of project milestones.

AI Engineer   **Jan 2021 - Jun 2022**
Zalo, VNG   **Hanoi, Vietnam · On-site**
- Developed text normalization modules for speech synthesis and speech recognition using a hybrid method: rule-based, language modeling, and deep learning (seq2seq) model.
- Developed and optimized services deployed on K8S environments, resulting in a 15% reduction in error rates.

Data Scientist   **Mar 2019 - Jan 2021**
Zalo, VNG   **Hanoi, Vietnam · On-site**
- Developed and applied speech products (ASR, TTS) to the Zalo ecosystem and KiKi virtual assistant, collaborating with a team of ten members.
- Optimized the language model to increase ASR's performance by up to 12% on products using ASR in the Zalo ecosystem (KiKi, LabanKey, ZingMP3, Zalo).
- Collected, analyzed, and cleaned data from various sources to support team tasks, leading to an increase in model accuracy for training AI tasks.

NLP Developer   **Jan 2018 - Mar 2019**
VCCORP   **Hanoi, Vietnam · On-site**
- The first full-time position while still a 4th-year student, demonstrating strong commitment and ability to balance work and study.
- Developed a semi-AI system to prevent spam comments on e-newspapers, utilizing Word2Vec for word similarity and classification models, significantly reducing manual labor.
- Deployed AI services to servers using RESTful APIs, facilitating seamless integration with the CMS system.
- Conducted experiments with Apache Hadoop and Spark for efficient data storage and computing, enhancing data processing capabilities.

SKILLS

Domain Knowledge: Large Language Model (LLM), Retrieval-Augmented Generation (RAG), Natural Language Processing (NLP), Information Retrieval, Data Crawling, Data Mining, Back-end Operations, System Design

Programming Languages: Python, Java, C/C++, Bash, HTML/CSS/JS

Tools & Frameworks: Git, Docker, MySQL, Kafka, Redis, Postgres (PGVector), Milvus, Pandas, Scikit-learn, Transformers, FastAPI

Soft-skills: Demonstrated strength in problem-solving and decision-making, coupled with a proactive initiative and a penchant for creativity and innovation.

PROJECTS

Luyện Code (LCOJ) · Personal 🔗   **Jan 2018 - Present**
- Developed and maintained an online coding challenge platform similar to Codeforces, serving over 10,000 Vietnamese students and offering AI-related IT services.
- Utilize a diverse tech stack including Python, PHP, HTML/CSS/JS, WordPress, and Docker for full-stack development and deployment.
- Manage self-hosting and blogging functionalities to ensure a seamless and reliable user experience.
- Enhance platform features to support educational needs and improve user engagement.

Simple RAG · Personal 🔗   **Apr 2024 - Apr 2024**
- Designed and implemented a RAG pipeline with a clear and efficient code structure, leveraging several advanced technologies:
  - **PDF Extraction**: Utilized Poppler for robust and accurate extraction of text from PDFs.
  - **Retrieval**: Employed BM25 algorithm for effective information retrieval.
  - **Reranking**: Integrated cross-encoder models to rerank retrieved results, enhancing relevance and accuracy.
  - **Self-Hosted LLM**: Deployed Ollama for language model generation, ensuring data privacy and control.

EDUCATION

Master's degree, Information Science   **Oct 2020 - Dec 2022**

Japan Advanced Institute of Science and Technology

Pursuing a master's degree abroad through a prestigious scholarship program.

Engineer's degree. Computer Software Engineering   **Jan 2014 - Dec 2019**

Electric Power University   **GPA:** 3.3

Commencing internship during second year while active in ACM/Olympic teams for three years.

CERTIFICATIONS

Shopee Code League 2021 - Participant 🔗 by **Shopee**   **May 2021**

Vietnamese Standardized Test of English Proficiency (VSTEP) - B2 🔗 by **Hanoi University**   **Apr 2021**

LANGUAGES

Vietnamese (Native proficiency) • **English** (Professional working proficiency)

PUBLICATIONS

Miko Team: Deep Learning Approach for Legal Question Answering in ALQAC 2022 🔗 on **International Conference on Knowledge and Systems Engineering (KSE)**   **Sep 2023**

Technical report for 1st rank in Task 01 (Document Retrieval) at ALQAC 2023

Development of Zalo Vietnamese Conversational Speech Recognition for VLSP 2019 🔗 on **VLSP**   **Nov 2019**

Technical report for Second prize, Speech-to-Text Shared Task VLSP 2019

AWARDS

First prize, Automated Legal Question Answering Competition by **International Conference on Knowledge and Systems Engineering**   **Oct 2022**

Second prize, Speech-to-Text Shared Task VLSP 2019 by **VLSP**   **Oct 2019**

Champion, AI4VN National Grand Finale Hackathon by **Topica, Kambria & MOST**   **Aug 2019**

Second Prize, ACM/ICPC VietNam National Programing Contest by **VAIP**   **Nov 2018**