VŨ VIỆT ANH

vietanhresearcher@gmail.com | github.com/VuVietAnh2309
linkedin.com/in/anh-vu-researcher | orcid.org/0009-0006-6977-5922



MỤC TIÊU NGHỀ NGHIỆP

Kỹ sư AI giàu nhiệt huyết với nền tảng vững chắc về Học máy (Machine Learning), Học sâu (Deep Learning) và Mô hình ngôn ngữ lớn (LLMs). Đam mê phát triển và triển khai các giải pháp AI sáng tạo, đồng thời đóng góp cho các nghiên cứu tiên tiến. Mong muốn kết nối nghiên cứu học thuật với ứng dụng thực tiễn trong môi trường năng động, với định hướng dài hạn là phát triển công nghệ AI và tiếp tục theo đuổi học vấn ở bậc cao hơn.

HỌC VẤN

Thạc sĩ Khoa học Máy tính – *Đại học Bách khoa Hà Nội*, Hà Nội, Việt Nam — 06/2025 – Hiện tại
CPA hiện tại: 3.71/4.0

Cử nhân Khoa học Máy tính (Hệ Tài năng) – *Đại học Bách khoa Hà Nội*, Hà Nội, Việt Nam — 10/2020 – 07/2024
CPA: 3.48/4.0

KINH NGHIỆM LÀM VIỆC

Prep Education, *Kỹ sư AI* | Hà Nội — 10/2025 – Hiện tại

Dự án: Hệ thống chấm điểm bài luận tự động (AES) cho phần Writing và Speaking

- Phát triển hệ thống chấm điểm tự động (AES) ứng dụng AI để đánh giá bài thi IELTS Writing Task 1 & 2 và phần thi Speaking với độ chính xác cao.
- AES CHO SPEAKING: Thiết kế và triển khai hệ thống chấm điểm phần thi Speaking sử dụng các mô hình Học máy tích hợp với ba hệ thống con chuyên biệt: APA (Đánh giá phát âm tự động) cho độ chính xác phát âm và độ trôi chảy, GEC (Sửa lỗi ngữ pháp) phát hiện và sửa lỗi ngữ pháp, và Dify để phân tích độ phong phú từ vựng. Đạt được các chỉ số: Accuracy $\pm 0.5 \sim 82\%$, QWK và PCC $\sim 90\%$.
- AES CHO WRITING: Thiết kế và huấn luyện các mô hình học sâu cho bài toán chấm điểm cross-prompt: PANN (mô hình Cross-Prompt dựa trên Deberta-V3) với 189M tham số, và mô hình T5 sau lượng tử hóa (quantized) tối ưu còn 120M tham số. Kết quả: Accuracy $\pm 0.5 \sim 85\%$ (Wt1) và $\sim 87\%$ (Wt2); Accuracy $\pm 1 \sim 98\%$.
- Triển khai pipeline MLOps end-to-end: huấn luyện mô hình $\rightarrow$ đóng gói Docker $\rightarrow$ tự động hóa CI/CD bằng GitLab $\rightarrow$ triển khai bằng ArgoCD lên cụm Kubernetes production.
- Tối ưu việc phục vụ mô hình bằng kỹ thuật dynamic batching và xếp hàng request để xử lý suy luận thông lượng cao cho hàng nghìn học viên đánh giá đồng thời.
- Đưa hệ thống vào production, cho phép chấm điểm tự động theo thời gian thực với độ trễ thấp và độ sẵn sàng cao.

Dự án: Hệ thống sửa lỗi ngữ pháp & từ vựng tiếng Anh (ELG)

- Phát triển hệ thống sửa lỗi ngữ pháp tiếng Anh dạng lai (hybrid), kết hợp phương pháp dựa trên luật (rule-based) với Mô hình ngôn ngữ lớn (LLMs) để nâng cao độ chính xác.
- Triển khai phát hiện và sửa lỗi toàn diện cho nhiều khía cạnh ngôn ngữ: lỗi ngữ pháp, dùng từ sai, lỗi chính tả, lỗi collocation, lỗi sử dụng thành ngữ, độ phù hợp về từ vựng, từ nối câu, và lỗi cụm từ — giúp học viên nhận diện và sửa lỗi viết hiệu quả.
- Thiết kế các module rule-based cho việc khớp mẫu xác định, kết hợp với suy luận từ LLM cho các cấu trúc ngữ pháp phức tạp.
- Đạt điểm F0.5 khoảng 60% trên tập kiểm thử riêng, cân bằng giữa precision và recall, ưu tiên precision để giảm thiểu sửa lỗi sai.
- Xây dựng hạ tầng phục vụ chuẩn production theo best practice MLOps: đóng gói Docker, GitLab CI/CD cho kiểm thử và triển khai tự động, ArgoCD cho điều phối Kubernetes.
- Triển khai chiến lược batching và caching request hiệu quả để tối ưu độ trễ suy luận của LLM, giảm chi phí tính toán đồng thời duy trì thông lượng cao.
- Triển khai hệ thống phục vụ hàng nghìn học viên mỗi ngày, hỗ trợ cải thiện kỹ năng viết qua việc sửa lỗi ngữ pháp và từ vựng theo thời gian thực.

Tập đoàn Công nghệ Vietsens, *Kỹ sư AI* | Hà Nội 10/2024 – 09/2025

Dự án: Trợ lý AI Y tế cho Quản lý Bệnh viện

- Phát triển giải pháp tích hợp AI trong lĩnh vực y tế, được triển khai trong Phần mềm Quản lý Bệnh viện Quốc gia, đặc biệt áp dụng tại Bệnh viện Bạch Mai.
- Sử dụng các kỹ thuật Học máy và Xử lý ngôn ngữ tự nhiên (NLP) tiên tiến, bao gồm Mô hình ngôn ngữ lớn (LLMs), để hỗ trợ y tế theo thời gian thực.
- Thiết kế và triển khai các module phân tích dữ liệu bệnh nhân, hỗ trợ quản lý ca bệnh và hỗ trợ chẩn đoán sơ bộ.
- Xây dựng các AI agent tương tác hỗ trợ bác sĩ và nhân viên y tế trong việc khai thác bệnh sử, kê đơn và chẩn đoán.
- Đảm bảo tích hợp liền mạch với hệ thống thông tin bệnh viện (HIS) hiện có nhằm nâng cao hiệu quả vận hành và chăm sóc bệnh nhân.

Dự án: Hệ thống gợi ý đơn thuốc và phát hiện xung đột thuốc

- Phát triển hệ thống dự đoán hỗ trợ bác sĩ kê đơn thuốc dựa trên thông tin bệnh nhân.
- Sử dụng các kỹ thuật phân tích dữ liệu để tiền xử lý và trích xuất đặc trưng từ hồ sơ bệnh án điện tử.
- Áp dụng các mô hình phân loại học máy truyền thống (Decision Trees, Random Forests, SVM) và thử nghiệm các bộ phân loại nâng cao để cải thiện độ chính xác.
- Tích hợp Mô hình ngôn ngữ lớn (LLMs) để đánh giá đơn thuốc, phát hiện trùng lặp hoạt chất và xác định khả năng tương tác thuốc–thuốc.
- Cung cấp công cụ hỗ trợ ra quyết định, giúp nhân viên y tế giảm sai sót khi kê đơn và nâng cao an toàn cho bệnh nhân.

Công ty Hệ thống Thông tin FPT, *Thực tập sinh Kỹ sư AI* | Hà Nội 08/2023 – 05/2024

Dự án: Sinh chú thích ảnh (Image Captioning)

- Phát triển hệ thống sinh chú thích cho ảnh sử dụng các mô hình học sâu tiên tiến.
- Xử lý dữ liệu hình ảnh để sinh các chú thích mô tả phản ánh chính xác nội dung của ảnh.
- Sử dụng cụ thể các mô hình tiên tiến như Transformer và Mạng nơ-ron tích chập (CNN) tích hợp cơ chế attention để nâng cao chất lượng sinh chú thích.
- Dự án nhằm xây dựng giải pháp mô tả ảnh tự động hiệu quả, ứng dụng các kỹ thuật mới nhất trong lĩnh vực thị giác máy tính và xử lý ngôn ngữ tự nhiên.

Dự án: OCR cho nhận dạng Báo cáo Tài chính

- Phát triển hệ thống OCR để nhận dạng và trích xuất thông tin quan trọng từ báo cáo tài chính của các ngân hàng.
- Áp dụng các mô hình OCR tiên tiến như PaddleOCR, Tesseract và TrOCR (OCR dựa trên Transformer) tích hợp NLP để cấu trúc hóa dữ liệu.
- Sử dụng LayoutLM / LayoutLMv3 để hiểu bố cục tài liệu, cải thiện việc trích xuất bảng, hình và các cặp khóa–giá trị.
- Triển khai tiền xử lý ảnh tài liệu (chỉnh nghiêng, nhị phân hóa, khử nhiễu) để tăng độ chính xác phát hiện và nhận dạng văn bản.
- Xây dựng pipeline hậu xử lý kết hợp Regex + Nhận dạng thực thể có tên (NER) để chuẩn hóa thông tin trích xuất.
- Triển khai mô hình như một phần của pipeline AI hỗ trợ các nhà phân tích dữ liệu tài chính.

Đại học Bách khoa Hà Nội, *Sinh viên Khoa học Máy tính* | Hà Nội 09/2022 – 09/2024

Dự án: Thích ứng với biến đổi khí hậu thông qua cải thiện dự báo thời tiết cực đoan (WIDS Datathon 2023 – Kaggle)

- Dự án phát triển mô hình dự đoán biến đổi khí hậu dựa trên các mô hình học máy.
- Xử lý dữ liệu dạng chuỗi thời gian (time-series) và áp dụng các kỹ thuật ensemble để nâng cao hiệu năng của mô hình.
- Sử dụng cụ thể các mô hình dự đoán mới nhất như Gradient Boosting, Light Gradient Boosting và CatBoost, kết hợp chúng với nhau.
- Dự án thuộc cuộc thi diễn ra cuối năm 2022 trên nền tảng Kaggle với mục tiêu dự đoán biến đổi khí hậu.

Dự án: Xác minh người nói tiếng Việt (Cuộc thi VLSP 2021)

- Dự án phát triển hệ thống xác minh người nói tiếng Việt sử dụng các mô hình học sâu.
- Xử lý các mẫu giọng nói để tạo ra đặc trưng âm học đặc thù của từng người nói phục vụ đăng ký và xác minh.
- Sử dụng cụ thể các mô hình như ECAPA_TDNN, ECAPA_CNN_TDNN và RawNet3, đồng thời thử nghiệm các hàm mất mát như triplet, softmax và proto.
- Dự án thuộc cuộc thi VLSP với chủ đề xác minh người nói, mục tiêu xác định chính xác hai mẫu giọng nói có thuộc cùng một người hay không.

Dự án: Phân tích cảm xúc theo khía cạnh (ABSA – NLP LAB)

- Xây dựng hệ thống phân tích cảm xúc theo khía cạnh cho tiếng Việt sử dụng SA-Transformer.
- Kết hợp các kỹ thuật NLP như gán nhãn từ loại (POS tagging) và Biaffine Parsing để trích xuất cặp khía cạnh–cảm xúc.
- Thu thập và gán nhãn dữ liệu chuyên ngành để huấn luyện và đánh giá mô hình.
- Đạt được độ chính xác phân loại cảm xúc theo khía cạnh tốt hơn so với các phương pháp baseline.

KINH NGHIỆM NGHIÊN CỨU

Nghiên cứu sinh viên cùng TS. Kate Han và TS. Nguyễn Thanh – 03/2023 đến nay

Học máy & Học kết hợp (Ensemble Learning)

VISTA: Thuật toán Lựa chọn Ensemble được Hỗ trợ bởi Surrogate dựa trên Giải thuật Di truyền chiều dài thay đổi và LSTM trong Hệ thống Ensemble nhiều tầng – *2024 IEEE Congress on Evolutionary Computation (CEC)* (Đã công bố, 06/2024)

- Dự án tập trung phát triển phương pháp VISTA, kết hợp Giải thuật Di truyền chiều dài thay đổi (VLGA) với mô hình surrogate dựa trên LSTM để tối ưu việc lựa chọn ensemble trong Hệ thống Ensemble nhiều tầng (MLES).
- Khảo cứu các phương pháp ensemble learning hiện có cùng thuật toán tiến hóa có hỗ trợ surrogate (SAEA).
- Đề xuất phương pháp mới kết hợp VLGA với LSTM để chuyển đổi mã hóa chiều dài thay đổi thành biểu diễn có kích thước cố định phục vụ dự đoán fitness.
- Triển khai hệ thống ensemble nhiều tầng, trong đó mỗi tầng có một tập các bộ phân loại (EoC) được huấn luyện trên cả dữ liệu gốc và dự đoán từ tầng trước.
- Thực nghiệm trên 15 bộ dữ liệu phổ biến, chứng minh VISTA vượt trội so với các thuật toán benchmark.
- Phát triển và xác thực thành công phương pháp VISTA, cho thấy hiệu quả trong việc cải thiện hiệu năng của MLES.

Khảo cứu các phương pháp Imputation

- Khảo cứu các phương pháp Imputation trên kho dữ liệu UCI Machine Learning Repository (15 bộ dữ liệu).
- Triển khai một số phương pháp Imputation (SimpleImputer, KNN, MICE, GINN, ...) cho dữ liệu thiếu và đánh giá trong bài toán phân loại.

Nghiên cứu sinh viên cùng PGS. Phạm Văn Hải và TS. Phạm Danh Tuyên – 09/2023 đến nay

Thị giác máy tính & Học sâu

Nâng cao Nhận diện Biểu cảm Khuôn mặt với CNN nhẹ dựa trên cơ chế Attention – *Journal of Ambient Intelligence and Humanized Computing* (Đã công bố, 08/2026)

- Thiết kế mô hình học sâu nhẹ cho nhận diện biểu cảm khuôn mặt, tối ưu để triển khai trên thiết bị biên (edge) có tài nguyên tính toán hạn chế.
- Sử dụng backbone MobileNetV2 cắt ngắn (truncated) để trích xuất đặc trưng hiệu quả từ ảnh khuôn mặt.
- Đề xuất khối Patch Extraction để chia feature map thành các vùng không chồng lấn, giúp mô hình tập trung vào đặc trưng cục bộ ngay cả khi bị che khuất hoặc thay đổi tư thế đầu.
- Tích hợp Attention Classifier với cơ chế self-attention nhằm tăng cường khả năng phân biệt đặc trưng và độ chính xác phân loại.
- Đạt kết quả cạnh tranh trên các tập dữ liệu chuẩn (RAF-DB, FER2013, FERPlus) cùng một bộ dữ liệu thực tế tự thu thập có yếu tố che khuất và biến đổi tư thế.

Nghiên cứu sinh viên cùng TS. Đinh Phú Hùng – 2026 đến nay

Hợp nhất ảnh y tế đa phương thức

CoRe-Fusion: Tái tổ hợp có kiểm soát cho hợp nhất ảnh y tế đa phương thức không cần huấn luyện (Đang phản biện)

- Hợp nhất không cần huấn luyện giữa ảnh chức năng (PET, SPECT) và ảnh giải phẫu (MRI, CT); xây dựng bộ dữ liệu não đồng đăng ký và một tập PET–CT bên ngoài.
- Đánh giá 11 phương pháp hiện đại trên cùng bộ chỉ số khách quan, công bố thành thư viện fusion-metrics.

PHẦN MỀM MÃ NGUỒN MỞ

fusion-metrics – Bộ chỉ số đánh giá chất lượng dung hợp ảnh, trung thành với MATLAB (Tác giả & người duy trì, 2026)

- Thư viện Python gồm 35 chỉ số đánh giá dung hợp ảnh ($Q^{AB/F}$, MI, SSIM, SD, NCIE, FMI, FSIM, VIF/VIFF, SCD, PSNR, CC, SF, entropy, average gradient), dịch từng dòng từ các bản cài đặt MATLAB tham chiếu.
- Kiểm chứng từng giá trị với file .m gốc: độ lệch ở mức sai số làm tròn ($\leq$ 1e-12) trên mọi chỉ số, kể cả các chi tiết riêng của MATLAB (căn tâm conv2 'same', bão hoà uint8, làm mảnh cạnh của edge(), thứ tự cộng).
- Phát hành trên PyPI (`pip install fusion-metrics`), có công cụ dòng lệnh để chấm cả một phương pháp, CI trên Python 3.10–3.13 và DOI Zenodo (10.5281/zenodo.24536175).
- github.com/VuVietAnh2309/fusion-metrics · pypi.org/project/fusion-metrics

KỸ NĂNG CHUYÊN MÔN

Ngôn ngữ lập trình	Python, Bash, C++
Framework Học máy	Keras, PyTorch
Thị giác máy tính	Phát hiện đối tượng, Phân vùng ảnh, Các thuật toán xử lý ảnh
Xử lý ngôn ngữ tự nhiên	Mô hình ngôn ngữ lớn (LLMs), các kỹ thuật NLP, Hệ thống Hỏi–Đáp (QA)
OCR & Xử lý tài liệu	PaddleOCR, Tesseract, TrOCR, LayoutLM, LayoutLMv3
MLOps & Triển khai mô hình	Docker, Kubernetes, GitLab CI/CD, ArgoCD, Tối ưu phục vụ mô hình, Dynamic batching, Chiến lược caching, Giám sát production
Lĩnh vực chuyên sâu	Học sâu, Thị giác máy tính, Mô hình ngôn ngữ lớn, Ứng dụng AI trong Y tế, Giáo dục

THÀNH TÍCH

- **Giải Nhì – VLSP 2026 Competition Award (Top 1 điểm Private Test)** (VLSP, 2026)

  Shared task Numerical Reasoning Question Answering (ViTNumChart) – Hỏi đáp tài chính (FinQA) với mô hình đa phương thức (VLM)

- **Giải Nhì – WIDS DATATHON 2023** (Kaggle, 2023)

  Thích ứng với biến đổi khí hậu thông qua cải thiện dự báo thời tiết cực đoan

NGOẠI NGỮ

Tiếng Anh	VSTEP B2 (25/03/2025). Có khả năng đọc hiểu tài liệu tiếng Anh tốt. Giao tiếp ở mức cơ bản.

CÔNG BỐ KHOA HỌC

- **Viet Anh Vu**, Gia Khanh Pham, Quy Nam Tran, Son Tung Do, Danh Tuyen Pham & Van Hai Pham. 2026. *"Enhancing facial expression recognition with lightweight attention-based CNN" – Journal of Ambient Intelligence and Humanized Computing*, Springer, công bố trực tuyến ngày 10/08/2026. DOI: 10.1007/s12652-026-05116-0.
- Kate Han, Truong Thanh Nguyen, **Viet Anh Vu**, Alan Wee-Chung Liew, Truong Dang & Tien Thanh Nguyen. 2024. *"VISTA: A Variable Length Genetic Algorithm and LSTM-Based Surrogate Assisted Ensemble Selection algorithm in Multiple Layers Ensemble System" – 2024 IEEE Congress on Evolutionary Computation (CEC)*, Yokohama, Nhật Bản, 30/06 – 05/07/2024, tr. 1–9. DOI: 10.1109/CEC60901.2024.10612029.

CHỨNG CHỈ

IBM Python for Data Science	Đã hoàn thành
Kaggle – Intro to Machine Learning	Đã hoàn thành
Kaggle – Computer Vision	Đã hoàn thành