

Đỗ Mạnh Hải

✉ haimanhdo1999@gmail.com | ☎ (+84)949109533 | [in linkedin.com/in/dsqt1999](https://www.linkedin.com/in/dsqt1999)
Hà Nội, Việt Nam

MỤC TIÊU NGHỀ NGHIỆP

Luôn nỗ lực học hỏi, phát triển và trau dồi kiến thức, công nghệ mới nhất trong lĩnh vực AI, tận dụng chuyên môn của mình để phát triển

HỌC VẤN

Đại học Bách khoa Hà Nội (HUST)

Kỹ sư Kỹ thuật Điều khiển và Tự động hóa;

Chuyên ngành: Kỹ thuật điều khiển và hệ thống thông minh

Hà Nội, Việt Nam

08/2017 – 04/2022

Trường THPT Chuyên Vĩnh Phúc

Chuyên Tin học;

Vĩnh Phúc, Việt Nam

08/2014 – 06/2017

Á khôi Kỳ thi học sinh giỏi Tin học cấp tỉnh lớp 11 và 12;

Kỳ thi THPT Quốc gia: **28.85/30**

KINH NGHIỆM LÀM VIỆC

Oxiitek

Kỹ sư AI

Hà Nội, Việt Nam

08/2024 – Hiện tại

- Nền tảng Tự động hóa Doanh nghiệp:** Thiết kế và triển khai các giải pháp tự động hóa toàn diện sử dụng n8n, thực hiện quy trình CRM tự động, quy trình phê duyệt, hệ thống thu thập phản hồi khách hàng và tạo task Jira liên mạch cho đội phát triển.
- AI Voice Bot cho Loa thông minh:** Phát triển voice bot thông minh dựa trên agent cho phép tương tác ngôn ngữ tự nhiên để quản lý và điều khiển thiết bị nhà thông minh với khả năng AI đối thoại tiên tiến.
- VoiceQA - Hệ thống Đánh giá Chất lượng Cuộc gọi AI:** Thiết kế và phát triển hệ thống đánh giá chất lượng cuộc gọi toàn diện được hỗ trợ bởi AI sử dụng Gemini 2.5 Flash để phân tích thông minh các cuộc gọi dịch vụ khách hàng. Triển khai RESTful API với FastAPI, tích hợp AWS S3 để lưu trữ âm thanh, cơ sở dữ liệu vector FAISS cho tìm kiếm ngữ nghĩa, và AWS Transcribe để xử lý speech-to-text.
- MCP cho Tích hợp IoT:** Phát triển server MCP (Model Context Protocol) toàn diện cho tích hợp hệ sinh thái API văn phòng, cho phép điều khiển và quản lý thiết bị IoT liên mạch thông qua LLM Agents với khả năng phản hồi thời gian thực.
- Nền tảng Tự động hóa Dịch vụ Khách hàng:** Xây dựng nền tảng tự động hóa end-to-end với tính năng chuyển đổi speech-to-text, tóm tắt cuộc hội thoại thông minh và tạo ticket hỗ trợ tự động với phân tích ngữ cảnh.
- Giải pháp Xử lý Tài liệu OCR:** Triển khai các giải pháp Nhận dạng Ký tự Quang học sẵn sàng sản xuất để xử lý tài liệu, thực hiện che giấu thông tin cá nhân bảo vệ quyền riêng tư và đếm nhãn sản phẩm tự động trong hình ảnh triển lãm.
- AWS Data Pipeline và Nền tảng Phân tích:** Thiết kế và triển khai kiến trúc data pipeline ETL toàn diện quản lý luồng dữ liệu từ DynamoDB đến Data Lake và lưu trữ S3. Điều phối quy trình xử lý dữ liệu tự động sử dụng AWS Glue với cronjobs được lên lịch để đồng bộ dữ liệu thời gian thực.

Job3s

Kỹ sư AI

Hà Nội, Việt Nam

04/2023 – 08/2024

- Hệ thống Tạo Nội dung LLM:** Phát triển hệ thống tạo nội dung thông minh sử dụng Large Language Models để tạo mô tả công việc tự động và tối ưu hóa CV, cải thiện chất lượng và tính nhất quán của nội dung tuyển dụng.
- Pipeline Web Scraping Đa trang:** Phát triển cơ sở hạ tầng data pipeline mạnh mẽ cho web scraping đa trang sử dụng BeautifulSoup, Scrapy và Selenium, xử lý và lưu trữ datasets quy mô lớn lên AWS S3 với xác thực dữ liệu tự động.
- Engine Gợi ý Nghề nghiệp AI:** Tạo engine gợi ý nghề nghiệp được hỗ trợ bởi AI sử dụng Elasticsearch để ghép cặp ứng viên-công việc thông minh và thuật toán dự đoán lương, nâng cao đáng kể hiệu quả và độ chính xác của quy trình tuyển dụng.
- Hệ thống Phân tích CV:** Xây dựng hệ thống phân tích CV toàn diện trích xuất 20+ trường có cấu trúc (thông tin cá nhân, học vấn, kinh nghiệm, kỹ năng) thông qua kỹ thuật xử lý PDF và OCR hybrid sử dụng VietOCR để nhận dạng văn bản tiếng Việt.
- Ghép cặp Tương đồng Ngữ nghĩa:** Triển khai hệ thống ghép cặp tương đồng ngữ nghĩa tính toán mức độ liên quan CV-mô tả công việc sử dụng PhoBERT-base embeddings và thuật toán cosine similarity, đạt được xếp hạng ứng viên độ chính xác cao.

VinBigData, Tập đoàn Vingroup

Kỹ sư NLP

Hà Nội, Việt Nam

10/2022 – 04/2023

- **Dataset Phân tích Cảm xúc theo Khía cạnh:** Xây dựng dataset phân tích cảm xúc dựa trên cực tính khía cạnh quy mô lớn cho đánh giá khách hàng Vinpearl, triển khai pipeline gán nhãn dữ liệu toàn diện, tiền xử lý và đánh giá mô hình Vietnamese DeBERTa.
- **Nghiên cứu ABSA với Context-Aware Attention:** Tiến hành nghiên cứu tiên tiến về Aspect-Based Sentiment Analysis so sánh BERT, BERT-pair, DeBERTa và cơ chế Context-aware attention trên datasets cảm xúc tiếng Việt.
- **Chatbot Sản phẩm được hỗ trợ bởi GPT:** Thiết kế và tiền xử lý dataset thông tin sản phẩm Shopee 400K+ cho hệ thống chatbot được hỗ trợ bởi GPT, cho phép truy xuất thông tin sản phẩm hiệu quả và giải quyết truy vấn khách hàng.

Viện VinBigData

Học viên Đào tạo Kỹ sư AI

Hà Nội, Việt Nam
07/2022 – 10/2022

- **Chương trình Chuyên môn AI:** Hoàn thành chương trình chuyên môn AI chuyên sâu bao gồm machine learning tiên tiến, deep learning, computer vision và xử lý ngôn ngữ tự nhiên với các dự án thực tế trong ngành.
- **Phát triển Chuyên môn:** Nâng cao năng lực chuyên môn trong làm việc nhóm, lãnh đạo kỹ thuật và kỹ năng thuyết trình thông qua chương trình đào tạo toàn diện được tài trợ bởi doanh nghiệp.
- **Dự án AI Đa lĩnh vực:** Thực hiện nhiều dự án AI bao gồm phân đoạn và phân loại polyp y tế, phát hiện bạo lực trong video streams và hệ thống hỏi đáp thông minh, thể hiện tính linh hoạt trên các lĩnh vực AI.

Smart Technology for Industry (ST4I)

Kỹ sư Computer Vision - Thực tập sinh

Hà Nội, Việt Nam
05/2021 – 07/2021

- **Tối ưu hóa Template Matching:** Tối ưu hóa thuật toán template matching trên các nền tảng Python, C++ và C#, đạt được cải thiện đáng kể về tốc độ xử lý và độ chính xác cho các ứng dụng xử lý dữ liệu công nghiệp.

HUST

Trợ lý Nghiên cứu

Hà Nội, Việt Nam
05/2020 – 04/2022

- **Nghiên cứu ML/CV:** Tiến hành nghiên cứu tiên tiến trong machine learning, computer vision và phương pháp deep learning, đóng góp vào tiến bộ trong thuật toán phát hiện cạnh và hệ thống nhận dạng khuôn mặt/đối tượng.
- **Hệ thống Dự đoán SpO2:** Phát triển hệ thống dự đoán SpO2 thời gian thực không tiếp xúc sử dụng computer vision dựa trên camera và kỹ thuật machine learning cho ứng dụng giám sát sức khỏe từ xa.

DỰ ÁN

Hệ thống Phân loại Banner Quảng cáo được hỗ trợ bởi AI | TensorFlow, OpenCV, FastAPI, PostgreSQL 01/2025 – Hiện tại

- **Pipeline Computer Vision:** Thiết kế và phát triển hệ thống deep learning toàn diện để phân loại tự động banner quảng cáo từ nội dung video thang máy sử dụng kiến trúc ResNet50/101/152 và MobileNetV3Large với kỹ thuật transfer learning.
- **Xử lý Video Tiên tiến:** Triển khai trích xuất frame thông minh sử dụng OpenCV và FFmpeg với phát hiện thay đổi cảnh dựa trên SSIM, đạt được lựa chọn frame tối ưu để nhận dạng banner với độ chính xác 95%.
- **API Sẵn sàng Sản xuất:** Xây dựng backend FastAPI có thể mở rộng với xử lý bất đồng bộ, tích hợp PostgreSQL và triển khai Docker hỗ trợ phân tích video thời gian thực và ánh xạ chiến dịch để theo dõi hiệu quả quảng cáo.
- **Data Augmentation & Tối ưu hóa:** Thiết kế pipeline data augmentation tùy chỉnh bao gồm random resize, hiệu ứng ánh sáng và biến đổi đặc thù cảnh để mô phỏng điều kiện ánh sáng thang máy thực tế, cải thiện độ mạnh mẽ của mô hình 15%.

WrenAI - Nền tảng Natural Language to SQL Doanh nghiệp | Python, FastAPI, Qdrant, BigQuery, Docker 04/2025 – Hiện tại

- **Kiến trúc AI Đa tenant:** Dẫn dắt phát triển dịch vụ AI đa tenant cấp doanh nghiệp hỗ trợ xử lý dữ liệu cô lập, cấu hình LLM đặc thù tenant và xác thực API an toàn với cô lập dữ liệu hoàn toàn sử dụng collections cơ sở dữ liệu vector Qdrant.
- **Tích hợp Nguồn Dữ liệu Động:** Phát triển tích hợp BigQuery liền mạch với khám phá schema tự động, tạo SQL động và profiling dữ liệu thông minh sử dụng phân tích được hỗ trợ bởi AI cho kết nối data warehouse doanh nghiệp.
- **Tạo Query được Nâng cao bởi RAG:** Triển khai pipeline Retrieval-Augmented Generation tiên tiến chuyển đổi truy vấn ngôn ngữ tự nhiên thành SQL được tối ưu hóa với xử lý nhận thức ngữ cảnh, hỗ trợ nhiều nhà cung cấp LLM và đạt được độ chính xác truy vấn 90%.
- **Kiến trúc Microservices:** Thiết kế và triển khai hệ sinh thái microservices có thể mở rộng bao gồm dịch vụ AI, thành phần UI, data engine và launcher với điều phối Docker, hỗ trợ 100+ tenant đồng thời với thời gian phản hồi dưới giây.

Nền tảng Chatbot RAG Doanh nghiệp | Django, Python, Docker, RAG, LLM

2023 – 2024

- **Xử lý Tài liệu Thông minh:** Thiết kế hệ thống chatbot RAG (Retrieval-Augmented Generation) toàn diện cho phép tương tác ngôn ngữ tự nhiên với kho tài liệu doanh nghiệp. Triển khai upload tài liệu thông minh, tổ chức theo chủ đề và khả năng tìm kiếm nhanh trên các bộ sưu tập tài liệu lớn.
- **Hỗ trợ Đa ngôn ngữ:** Phát triển nền tảng AI đối thoại với xử lý ngôn ngữ tự nhiên tiên tiến hỗ trợ tiếng Việt và tiếng Anh, có tính năng phản hồi nhận thức ngữ cảnh và tóm tắt cuộc hội thoại thông minh để nâng cao trải nghiệm người dùng.

- **Kiến trúc Web Có thể Mở rộng:** Xây dựng ứng dụng web Django sẵn sàng sản xuất với containerization Docker, có tính năng xác thực người dùng, bảng quản trị để quản lý nội dung và giao diện web responsive được tối ưu hóa cho cả môi trường phát triển và sản xuất.
- **Tích hợp Doanh nghiệp:** Thiết kế kiến trúc modular hỗ trợ tích hợp dễ dàng với các hệ thống doanh nghiệp hiện có, pipeline triển khai tự động và tài liệu API toàn diện để áp dụng tổ chức liền mạch.

Bộ Tự động hóa Dịch vụ Khách hàng AI | *Python, NLP, FastAPI, PostgreSQL, Docker*

01/2025 – Hiện tại

- **Chatbot Hỗ trợ Khách hàng Thông minh:** Phát triển chatbot dịch vụ khách hàng được hỗ trợ bởi AI toàn diện sử dụng kỹ thuật NLP tiên tiến để giải quyết truy vấn tự động, phân loại ticket và định tuyến thông minh đến các kênh hỗ trợ phù hợp với tỷ lệ giải quyết liên hệ đầu tiên 85%.
- **Bot Thu thập Thông tin Hàng không:** Phát triển chatbot chuyên biệt cho ngành hàng không thu thập thông tin khai báo chuyến bay, xác thực tuân thủ quy định và tự động hóa quy trình nhập dữ liệu cho cơ quan hàng không với xác thực thời gian thực và xử lý lỗi.
- **Tạo Báo cáo PDF Tự động:** Xây dựng hệ thống thông minh để tạo báo cáo PDF toàn diện cho quy trình đăng ký bay drone nông nghiệp, bao gồm điền form tự động, kiểm tra tuân thủ quy định và tích hợp chữ ký số với cơ sở dữ liệu hàng không chính phủ.
- **Tích hợp Đa kênh:** Triển khai triển khai chatbot omnichannel hỗ trợ giao diện web, ứng dụng di động và tích hợp API với hệ thống CRM hiện có, cho phép trải nghiệm khách hàng liền mạch trên tất cả các điểm tiếp xúc.

NĂNG LỰC CÁ NHÂN

- **Ngôn ngữ Lập trình:** Python, C++, C#, JavaScript, HTML, CSS, SQL
- **Framework AI/ML:** TensorFlow/Keras, PyTorch, LangChain, LangIndex, Hugging Face Transformers, OpenAI API
- **Framework Web:** FastAPI, Django, Flask, RESTful APIs, Unicorn
- **Cơ sở Dữ liệu & Lưu trữ:** PostgreSQL, MySQL, MongoDB, FAISS Vector Database, Qdrant, AWS S3, ElasticSearch
- **Cloud & DevOps:** AWS (S3, Glue, Athena, QuickSight, Transcribe, DynamoDB), Docker, Docker Compose, Microservices
- **Chuyên môn AI/ML:** RAG (Retrieval-Augmented Generation), NLP, Computer Vision, OCR, Semantic Search, LLM Integration
- **Xử lý Dữ liệu:** ETL Pipelines, Data Lake Architecture, Business Intelligence, PDF Processing, Audio Processing
- **Quản lý Dự án:** Git, GitHub, GitLab, Trello, Confluence, Agile, CI/CD
- **Khác:** Linux, Postman, n8n Automation, IoT Integration, MCP (Model Context Protocol)