NGUYEN HUU KHANG

Phone Number: 0945 466 497
Email: nskhang1@gmail.com
LinkedIn / GitHub

OBJECTIVE

I aim to work in a challenging environment where I can meet with human pain points/problems and transform my AI experience into real-world applications that deliver real and meaningful value to clients, enhancing their satisfaction and making their lives/businesses more convenient.

TECHNICAL SKILL

Language	Vietnamese, English.
Base Knowledge	Machine/Deep learning, Computer Vision, Large Language Model, Multimodal Large Language Model, DSA.
Programming Languages	Python, SQL.
Frameworks/Libraries:	Pytorch, TensorFlow, Keras, Pytorch lightning, Transformers, OpenCV, NumPy, Matplotlib, Pandas, Flask, Dash, FastAPI,...
Tools	CVAT, DVC, Airflow, MLflow.

WORK EXPERIENCE

AI Engineer at GotIT May 2024 - Present

1. **ScanIT:**
   - Description: ScanIT empowers clients with post-purchase care by offering advanced bill-scanning capabilities and extracting key information, reducing promotion campaign operational costs by 30%.
   - My responsibility: Designed and developed an OCR engine and served as API to accurately scan and extract critical information from bills, enabling merchants and brands to efficiently manage scalable and reliable promotional campaigns. Supported a diverse range of clients across industries such as F&B, FMCG, and Beauty, including major brands like Heineken, Sharp, Samsung, and Toshiba, ensuring seamless integration and high-quality service delivery.
   - TechStack: AWS EC2, Docker, Docker Compose, FastAPI, Redis, Triton, GRPC, Pytorch, Airflow, MLFlow, Superset.
2. **GotIT Chatbot:**
   - Description: A chatbot designed to support internal business processes and enhance operational efficiency.
   - My responsibility: Implemented a chatbot leveraging the Retrieval-Augmented Generation (RAG) technique, seamlessly integrating it into Microsoft Teams to optimize internal workflows and elevate customer service quality. Reduced customer services costs and delivery accurately answer for customer.
   - TechStack: AWS EC2, AWS RDS, Redis, Transformers, Chainlit, Llama Index, Langfuse, FastAPI, Docker, Docker Compose.
3. **Multimodal Large Language Model:**
   - Description: Researching and preparing proposal to apply multimodal large language models to our ScanIT.
   - My responsibility: preparing data, and fine-tuning state-of-the-art models such as MiniCPM-V, InternVL, Monkey-OCR to benchmark their capabilities and costs when applied to Vietnamese data.

- TechStack: AWS EC2, Transformers, PyTorch.

AI Engineer at GMOz.com-RUNSYSTEM Sep 2021 - May 2024

1. **Text Recognition model:**
   - Description: Recognizing text line by line in images using a cross image-language model, aiding the business in digital transformation.
   - My responsibility: Handling Vietnamese and Japanese text, both printed and handwritten. Modeling and training deep learning models. Conducted error analysis and proposed solutions to enhance model accuracy and inference speed. Building data pipelines and automated training pipelines. Customizing models to meet specific customer requirements and use cases. This project achieved an average of 97% accuracy in Vietnamese and 91% accuracy in Japanese.
   - TechStack: PyTorch, TensorFlow, MLFlow.
2. **Table Reconstruction:**
   - Description: Reconstructing tables in document images into CSV/Excel files to help users easily transform image documents into digital formats.
   - My responsibility: Modeling deep learning model to reconstruct border tables from document images, combining this with OCR pipeline to produce output as CSV format. This project achieved over 90% accuracy on an internal dataset with 2 fps (running on RTX 2080).
   - TechStack: MLFlow, PyTorch, Docker.
3. **SmartOCR:**
   - Description: Product to transform freeform document image into digital text.
   - My responsibility: Develop a scalable, easy-to-maintain OCR API pipeline integrating multiple models, such as text detection, text recognition, image alignment, automatic document detection, auto-rotation. This solution is tailored for various clients, including banks, insurance companies, and government agencies, supporting both on-premise and cloud deployments. This product processes hundreds of thousands of requests per month, digitizing millions of document pages in both Vietnamese and Japanese. Its adoption has significantly contributed to business growth, driving substantial revenue and operational efficiency of client.
   - TechStack: Flask API, TensorFlow, PyTorch, ONNX, Docker.

EDUCATION

Bachelor of Science in Computer Science - GPA 3.2/4.0 2018 - 2022
University of Information Technology - VNU HCMC (UIT).