Dương Đình Hiếu

AI Engineer

CONTACT

Phone: 0353728206

Email: hieuai0305@gmail.com

Address: Hòa Lạc, Hà Nội

Github: https://github.com/HieuAI2005

PROJECT

Extract Key–Value from Vietnamese administrative document PDF file | 2025
Personal

- Fine-tuned RoBERTa for Vietnamese document NER (7 labels); delivered micro-F1 0.95 with a production-ready IDP pipeline (PyTorch/HF, AdamW, BIO)
- Designed the end-to-end pipeline (text normalization, BIO tagging, BPE tokenization, seqeval evaluation) and exported JSON outputs for an IDP workflow.

Methods & tools: Python, PyTorch, Hugging Face Transformers (RoBERTa), AdamW, linear warmup/decay scheduler, gradient clipping, early stopping

Document Intelligence system in multilingual electronic invoice processing| 2025
Personal

- Built an end-to-end pipeline for multilingual invoice understanding, from image preprocessing to structured JSON outputs.
- Implemented robust image preprocessing (de-skew, illumination correction, denoising, contrast/thresholding) to stabilize OCR and layout detection.
- Trained a Detectron2 layout/field detector and reached ~0.94 average precision on the hold-out set
- Reconstructed table structure with PaddleOCR TableCellsDetection and a custom GNN to infer cell adjacency, reading order, and header–value links.
- Leveraged Qwen 2.5 (1.5B) for multilingual text normalization, value inference, and post-processing checks.
- End-to-end throughput ≈30 seconds per page in production tests.

Methods & tools: OpenCV, Detectron2, PaddleOCR (TableCellsDetection), PyTorch (GNN), Qwen 2.5 1.5B, AdamW, data augmentation, early stopping, linear warmup/decay scheduler.

EDUCATION

FPT University | 2023-Now
Artificial Intelligence

- GPA: 3.1
- Expert Notebook in Kaggle