TOAN, NGUYEN VAN

Artificial Intelligence Engineer

(+84) 362305676 ⋄ nvantoan1203@gmail.com

linkedin.com/in/toan-data-861623272 ⋄ github.com/noaft

SUMMARY

Proficient in designing, building, and maintaining efficient data pipelines with solid expertise in Python and advanced data processing tools. Experienced in implementing scalable solutions that drive business insights using technologies like Apache Spark, Hadoop, and Apache Airflow. Skilled in SQL and various database management systems, including MySQL, PostgreSQL, and MongoDB, have strong background about deep learning and math. Committed to continuous learning and staying updated with the latest advancements in AI and data engineering. Known for strong problem-solving skills, a detail-oriented approach, and effective collaboration with cross-functional teams.

EDUCATION

University of Information Technology   *April 2021 - Present*
4rd year majoring in information technology

LANGUAGE PROFICIENCY

- English: Intermediate (Can read paper, documents and listen, basic listening and reading)

- Japanese: Intermediate (conversational and reading, basic listening and reading)

KNOWLEDGE

- **Mathematics:** Intermediate (Proficient in calculus, linear algebra, and statistics for data analysis and machine learning)

- **Artificial Intelligence:** Knowledgeable in Natural Language Processing (NLP) and Computer Vision (CV), including working with deep learning models for text and image understanding.

TECHNICAL SKILLS

Language	Python, C++, HTML, CSS, Javascript
Database	MySQL, SQL server, Mongodb
Vector Database	Faiss, Milvus
Developer Tools	Git, Kafka, Hadoop, Spark, SSIS, SSAS
Frame Work	FaspAPI, Pytorch, Tensorflow
Operator System	Window, Ubuntu

EXPERIENCE

Freelance AI Engineer: Development of AI Applications for IoT Devices:

- **Socker/TCB Integration:** Designed and implemented Socket connections to seamlessly interact with IoT devices, ensuring real-time data retrieval, transmission, and secure storage..

- **AI Model Development**: Engineered and optimized a logistics linear model tailored for smell classification, addressing domain-specific challenges and improving predictive accuracy

- **Model Training Pipeline:** Conducted in-depth research and analysis on collected data. Applied preprocessing techniques such as Principal Component Analysis (PCA) and data normalization to prepare high-quality inputs, enabling robust and efficient model training.

- **Exploratory Data Analysis:**
  - **Preprocessing:** Standardize the dataset.
  - **Dimensionality Reduction:** Apply PCA to reduce dimensionality to 3D.
- **Full-Stack Web Development:** Developed and deployed a web-based solution encompassing both backend and frontend components. The system integrates IoT device data collection, secure storage, model creation, data labeling workflows, and model training, providing a cohesive and user-friendly interface.

RAG Image and Text Web:

- **Collect data:** Leveraging OpenCV to extract frames from videos for analysis and processing.
- **Pretrained Model:** Integrating the BEiT3 model to enhance website capabilities and support efficient data processing .
- **Vector database:** Utilizing Milvus to efficiently store and perform complex queries on vector data.
- **Web Design:** Developing a professional and user-friendly frontend interface to ensure optimal user experience.

CNN Classification for Feature Engineering in Malware Detection

- **Data Collection:** Gather malware datasets from online sources.
- **Prediction:** Train the model using the collected data and perform predictions.
- **GitHub:** `https://github.com/noaft/Federated_and_CNN_for_classification`

Using Yolov8 for real-time detection of people not wearing helmets:

- **Data Collection:** Sourced comprehensive datasets from various internet resources.
- **Model Training**: Developed a YOLOv8 model using Python.
- **Video Processing:** Employed OpenCV to split frames from virtual camera feeds and input data into Kafka.
- **System Design:** Architected a real-time system utilizing Kafka for data ingestion and Spark for processing.
- **Classification and Storage:** Applied the pre-trained YOLOv8 model for classification within Spark modules and stored data, including images of non-helmeted individuals and traffic vehicle counts, in SQL databases
- **GitHub:** `https://github.com/noaft/Detected-motorcyclist-not-wearing-a-helmet`