

# Huy Pham

---

As a Data Scientist with Natural Language Processing and Database Management experience and an MSc in Data Science and AI, I bring a strong foundation in machine learning, data analytics, and programming to any organization. I have demonstrated a commitment to excellence and a passion for continuous learning and innovation.

## Technical Skills

---

**Programming:** Python, SQL, VBA, PHP

**Machine Learning:** LLM, LangChain, TensorFlow, PyTorch, scikit-learn, Flair, spaCy, ...

**Data Collection:** Qualtrics, REDCap

**Website Development:** PHP, WordPress, HTML, CSS

**Others:** Azure, Google Colab, GitLab, Hypothesis Testing, MySQL, PySpark

## Working Experience

---

### **Dec 2023 – Current. Data Coordinator, University of Cambridge, UK**

Designed and developed programs for data validation, cleaning, and re-formatting, utilizing SQL and Python. Developed secure electronic data entry forms based on Case Report Forms (CRFs), ensuring data integrity and compliance with regulatory requirements. Collaborated with public health researchers and statisticians to analyse data, facilitate seamless data flow and resolved data-related issues. Built and maintained project websites, ensuring up-to-date content and an engaging user experience. Collected, processed, and stored project data in accordance with General Data Protection Regulations (GDPR) and Data Protection Act.

### **Nov 2022 – May 2023 (FTC). Data Scientist – KTP Associate, Aston University & Folding Space company, UK**

Built custom training datasets for Named Entity Recognition (NER), through a combination of automated and manual annotation processes. Managed the NER training processes, curating a collection of domain-specific models and repackaging them for redistribution in Folding Space's products. Collected feedback about performance of the NER models "in the field" and refined the models in response to the feedback.

### **Oct 2020 – Aug 2021. Research Assistant, Department of Toxicology, Institute of Environment and Resources, Vietnam**

Conducted data-driven research on water quality, phytoplankton density, and plastic pollution in Ho Chi Minh City's rivers. Designed innovative experiments to analyse the feeding behaviour of aquatic organisms in response to microplastic exposure and developed a scalable protocol for extracting microplastics from river sediment. This work highlighted my role as a data scientist contributing to understanding environmental dynamics and plastic pollution impact.

## Education

---

### **09/2021 – 09/2022, MSc in Data Science and AI -2.1 (Merit), University of Liverpool, UK.**

- Relevant courses: Data Mining and Visualisation, Computational Intelligence, Machine Learning and Optimisation, Database and Information System, Applied AI.

### **09/2016 – 09/2020, BSc in Aquatic Resources Management – 2:1, International University, Vietnam.**

- Relevant courses: Biostatistics, Calculus, Physics, Chemistry, Biology, Biochemistry.