# Nai Jui Yeh (Nary Yeh)

(+44) 7513 625747 | nary12321@gmail.com | https://www.linkedin.com/in/nary-yeh-1583b3213/

## Summary

An ambitious and passionate MSc Data Science student from the University of Nottingham with experience in Python programming, machine learning, and back-end development. I am seeking a graduate software engineer opportunity to leverage my skills in Python and AI as well as my planning and problem-solving skills to develop and deploy cutting-edge platforms.

## Education

**University of Nottingham** Sept. 2023 - Sept. 2024
MSc Data Science: distinction (predicted) Nottingham, UK

- **Coursework:** Machine Learning, Big Data Learning and Technologies, Statistical Inference, Time Series and Forecasting

**National Taipei University of Technology** Sept. 2018 - June 2023
Intelligent Automation Engineering: GPA: 3.99/4.0 Taipei, Taiwan

- **Coursework:** Data Structure, Algorithm, Computer Network, Database System, Machine Learning

## Work Experience

**AI Back-End Engineer** July 2023 - Dec. 2023
Genenet Technology (UK) Cambridge, UK

- Developed and deployed scalable **data pipelines** and **RESTful APIs** using **Python** for a bioinformatics analysis application
- Implemented comprehensive **unit tests** to validate system functionalities and ensure robustness
- Optimized system efficiency by 20% by building an asynchronous task queue architecture using **RabbitMQ** and **Redis**
- Deployed the full-stack application on **GCP**

**Deep Learning Intern** July 2022 - Aug. 2022
MediaTek Research Taipei, Taiwan

- Developed and deployed an **NLP** application in **Python** with **PyTorch** and **FastAPI**, adopted by multiple MediaTek business units
- Developed **MongoDB** schemas and implemented **RESTful API** for efficient data access
- Enhanced the functionality of the large language model by crafting effective **prompts** and input data
- Simplify backend system and machine learning model deployments using **Docker** containers

## Education Projects

**Big Data Approach to Improve Genetic Prediction in Alzheimer's Disease** Feb. 2024 - May 2024
*PySpark, MLlib, Databricks*

- Developed a scalable Big Data pipeline using **PySpark** and **MLlib** on **Databricks**
- Analyzed large-scale genetic sequence data (7 million features) for Alzheimer's Disease prediction

**Breast Cancer Treatment Response Prediction** Sept. 2023 - Dec. 2023
*Python, Machine Learning, Scikit-learn, Pandas, NumPy*

- Developed machine learning models using Python with **Scikit-learn**, **Pandas**, and **NumPy,** achieving a 30% improvement in disease treatment response prediction
- Optimized model performance through feature reduction techniques like PCA
- Collaborated with a multidisciplinary team to deliver a **data-driven** approach for improved disease treatment prediction