Resume

Yi-Hua Guo 郭益華

g11350006@thu.edu.tw
jerry7776112@gmail.com
Linkedin
Github

Education

02/2022-Present GPA: 4.12/4.30

Tunghai University, Taichung, Taiwan(東海大學)

Master of Engineering, Computer Science(資訊工程學系)

06/2016-06/2020 GPA: 3.17/4.30

Tunghai University, Taichung, Taiwan(東海大學)

Bachelor of Engineering, Environmental Science and Engineering(環境科學與工程學系)

Courses taken include:

- C++ Programming （C++程式設計）
- Decision Support System（決策支援系統）
- Internet of Things (IOT) for healthcare（照護物聯網）
- Information Assurance（資訊安保）
- Study and Implementation of Big Data and Industry Applications（大數據應用與產業案例實作）, Advanced Software Engineering Special Topic（高等軟體工程專題）
- Information Management（資訊管理）
- 5G Network Systems and Applications（5G 網路系統與應用）
- The Business Value Creation of Digital Technology（數位科技的商業價值創造）
- New Economy and Business Model Innovation（新經濟與創新營運）
- Deep Learning（深度學習）
- The Fundamentals and Applications of Data Science Modeling（資料科學建模的基礎及應用）

Certificates:

- TUV NORD - Sustainability Management Manger(企業永續管理師)
- 2023 AIS3
- 2021 企業數據競賽
- Career Essentials in Data Analysis by Microsoft and LinkedIn
- Career Essentials in Generative AI by Microsoft and LinkedIn
- AWS Academy Graduate - AWS Academy Cloud Foundations
- Great Learning - Bascis of EDA with Python

Experience

06/2023-Present Research Assistant, Secure Social Computing Laboratory

- Produced data engineering side projects, including building openweather ETL using AWS EC2 and S3 and deploying Python-Flask applications using Docker. Use Power BI to create a dynamic visual dashboard for pizza sales data.
  （製作資料工程 side project，包括使用 AWS EC2 和 S3 建置 openweather ETL 以及使用 Docker 部署 Python-Flask 應用程式。 使用 Power BI 為披薩銷售資料建立動態視覺化儀

表板。）

Skill Sets: AWS, Apache Airflow, Python, Python-Flask Framework, Makefile, Dockerfile, PostgreSQL, MongoDB, MS SQL, Power BI

- Use "Tryhackme" to learn Cyber Security and obtain certificates including "Introduction to Cybersecurity", "Pre Security", "Jr Penetration Tester" and more. The skills include Network Exploitation, Web Exploitation, Linux Exploitation, Windows Exploitation, Pentesting.
  （使用「Tryhackme」進修網路安全課程並獲得包括「網路安全入門」、「進階網路安全」、「初級滲透測試員」等證書。 技能包括網路利用、Web 利用、Linux 利用、Windows 利用、滲透測試。）

Skill Sets: Burp Suite, Metasploit, Nmap, Nessus, Hydra

02/2022-06/2023 Research Assistant, Secure Social Computing Laboratory

- Collect public data from social networking platforms for big data analysis and disinformation tracking.
  （從社群網路平台收集公開資料，進行大數據分析和假新聞追蹤。）
- For Cyber security, build a "Penetration Testing" and "Vulnerability Assessment" environment using VMware virtual machine, and use a second virtual machine, Kali Linux, as the attacking party for testing.
  （網路安全方面，使用 VMware 虛擬機器建立『滲透測試』和『漏洞評估』環境，並使用第二個虛擬機器 Kali Linux 作為攻擊方進行測試。）
- Paper review, paper writing, and presentation skills.
  （論文審查、論文撰寫和演講技巧、滲透測試技巧。）

Skill Sets: Virtual Machine, Penetration Testing Concepts, Vulnerability Assessment Concepts, Cyber Security Implementation, Big Data Analysis, Big Data Processing, Social Network Analysis, paper review, paper writing, presentation skills

01/2023-6/2023 Research Assistant-資安卓越中心規劃建置計畫, National Institute of Cyber Security(國家資通安全研究院）

- Based on the about 30 million pieces of public data of social networks were obtained, developed API & ETL pipeline.
  （利用取得的約 3,000 萬個社群網路公開資料，開發了 API 和 ETL 程序。）
- A social network analysis study was conducted using approximately 30 million pieces of public data.
  （使用大約 3000 萬個公開資料進行了社群網絡分析研究。）
- Based on the research questions, we visualized the social network for the social network message dissemination phenomenon study.
  （根據研究問題，我們將社群網路視覺化，進行社群網路訊息現象研究。）

Skill Sets: Big Data Analysis, Big Data Processing, Data Visualization, Social Network Analysis

07/2022-12/2022 Research Assistant-資安卓越中心規劃建置計畫, National Institute of Cyber Security(國家資通安全研究院）

- The VMware virtual machine build uses a combination of deep learning semantic similarity analysis and artificial intelligence to develop an online social network disinformation tracking API.
  （使用 VMware 虛擬機建構了結合深度學習語意相似性分析和人工智慧來開發線上社群網路可疑訊息追蹤 API。）
- Packages VMware for delivery as a VMDK file.

（將開發後的 VMware 打包為 VMDK 檔案。）

- The development process uses Git for version control and facilitates collaboration with other engineers.
  （開發過程中使用 Git 進行版本控制，方便與其他工程師的協作。）
- About 30 million social network public data were built into the PostgreSQL database, and design the database index to improve the search performance from the original 30 seconds per search to 0.77 seconds.
  （將約 3000 萬條社群網路公開資料建置在 PostgreSQL 資料庫中，並設計資料庫索引，將搜尋效能從原來的每次搜尋 30 秒提高到 0.77 秒。）
- Using Python Flask web application framework was combined with PostgreSQL to develop the API
  （使用 Python Flask 框架結合 PostgreSQL 開發 API。）
- About 6 million social network public data were built into the MongoDB database, and design the database index to improve the search performance.
  （將約 600 萬則社群網路公開資料建置到 MongoDB 資料庫中，並設計資料庫索引以提高搜尋效能。）
- Using Python Flask web application framework was combined with MongoDB to develop the API
  （使用 Python Flask 框架結合 MongoDB 開發 API。）

Skill Sets: Virtual Machine**,** Natural Language Processing, Deep Learning, API Development, SQL, Git, Team-Work

02/2022-6/2022 Teaching Assistant, Tunghai University, Computer Science Department

- Instructing discussion sessions in the course of Artificial Intelligence.
  （在人工智慧概論課程中指導學生與帶領討論。）

Skill Sets: Basic Artificial Intelligence Concepts, Basic Computer Science Concepts, Communication & Team-Work

06/2021-10/2021 Member, Workforce Development Agency, MOL, AI Big Data Enterprise Practice Application Class

- Learn basic concepts of Python and R programming, data processing skills, as well as data visualization and artificial intelligence machine learning concepts.
  （學習 Python 和 R 程式設計的基本概念、資料處理技能，以及資料視覺化和人工智慧機器學習概念。）
- Participated in the "2021 Data Station" competition, using the "PChome product purchase list" dataset provided by the organizer.
  （參加「2021 Data Station」企業數據競賽，使用主辦單位提供的「PChome 產品購買清單」資料集進行分析。）
- Based on the dataset, I processed and added new features to the data, and used third-party data to overlay it, and used Python machine learning algorithms such as decision trees and random forests to analyze the frequency of customers' purchases.
  （根據資料集，對資料進行了處理和新增特徵，並使用第三方資料對其進行疊加，最後使用決策樹、隨機森林等 Python 機器學習演算法來分析客戶的購買頻率。）

Skill Sets: Python Programming, R Programming, Data Processing Skills, Data Visualization, Data Analysis Based on Enterprise Data, Basic Artificial Intelligence Concepts

Projects	

- DockerTutorial(Zero to Hero)-NOTE
- ETL 語意相似度分析
- Python & Kafka 餐點訂單分散式系統設計
- SQL to Power BI Pizza 銷售資料視覺化
- 使用 pgAdmin 操作 Docker 上的 PostgreSQL
- 系統 CPU & Memory 監控 Flask & Docker 實作
- 利用 AWS 建立 Open Weather ETL 並整合至 Apache Airflow
- 資料模型設計實作
- AIS3-信用卡盜刷偵測
- 新經濟創與創新營運(創新提案)_穆斯林 SuperApp
- 數位科技的商業價值創造(創新提案)_虛擬生命園區平台

Computer Skills

Languages: Python, R, C++, Go
AI Library: Sentence-Transformers, Scikit-learn
Data Processing: Pandas, Numpy
ETL Tool: Airflow, SSIS
Data Visualization: PowerBI, Seaborn, Matplotlib, Plotly
Social Network Visualization: Gephi
Python Web Framework: Flask

SQL: MySQL, PostgreSQL, MongoDB
Virtual Machine: VirtualBox, VMware, Hyper-V
OS: Windows, Linux
Version Control: Git
Editor: VScode, Nano, Vim
Scripts: Bash
Cloud: AWS
Container: Docker, Kubernetes

Publications

- Tracking of Disinformation Sources Based on Online Social Media: Examining Pages and URLs with BFS Evolution | CISC2023 第三十三屆全國資訊安全會議
- 基於區塊鏈建立去中心化社群平台之共識演算法研析 | CISC2022 第三十二屆全國資訊安全會

Awards

Session C4 : Social Media Security-Best Session Award
Tracking of Disinformation Sources Based on Online Social Media: Examining Pages and URLs with BFS Evolution | CISC2023 第三十三屆全國資訊安全會議