

Waihong Woo, Seth

Aspiring AI/ML Researcher, Former Data Scientist (R&D)

Ipoh, Perak, MY / sethwoo555@gmail.com | +6012 6942981

linkedin.com/in/sethwoo / github.com/seth-woo | seth-woo.github.io



About Me

I am a PhD applicant and aspiring AI/ML researcher. Previously, I was a R&D Data Scientist in a sports analytics startup where I designed, trained and optimized DL architectures, with a focus on attention-based models. I hold a Master's degree in Data Science from the University of Malaya (UM) and a Bachelor's degree in Computer Science from Sunway University. For my PhD, I aim to specialize in World Models and Artificial General Intelligence. I am excited to be a part of the movement that builds the next frontier of foundation models that are capable of understanding how the world works through internal mental models, possess intrinsic motivation, can reason and plan complex actions and have persistent memory, all while remaining controllable and safe for real-world adoption.

Education

| | |
|---|---------------------|
| University of Malaya , Master of Data Science | Mar 2023 – Feb 2024 |
| <ul style="list-style-type: none">• Pass (with Distinction) (CGPA: 4.0 out of 4.0)• Master's Thesis: Bank Customer Churn Analysis using Machine Learning Methodologies• Coursework: Machine Learning, Numerical Optimization, Core Data Science & Analytics, Big Data Technologies, Research & Project-Based Learning, Distributed Systems | |
| Sunway University-Lancaster University , BSc (Hons) in Computer Science | Mar 2019 – Jul 2022 |
| <ul style="list-style-type: none">• First Class Honours (CGPA: 3.93 out of 4.0)• Bachelor's Thesis: Cyberbullying Conceptualization, Characterization and Detection in Social Media - A Systematic Literature Review• Coursework: Artificial & Computational Intelligence, Statistics, Core Computer Science Fundamentals, Digital Image Processing, Web Fundamentals, Object-Oriented Programming, Functional Programming, Capstone Project | |
| Sunway College , Cambridge GCE Advanced Level (A-Level) | Jul 2016 – Nov 2017 |
| <ul style="list-style-type: none">• Grade: A*A*A*• Subjects Taken: Physics, Chemistry, Mathematics, Further Mathematics | |

Projects

| | |
|--|----------------------|
| Multimodal Knowledge Retrieval Assistant (MKRA) (Ongoing) | Nov 2025 - Present |
| <ul style="list-style-type: none">• MKRA will combine pre-trained embedding models with a LangChain RAG pipeline, Qdrant-Docker stack for vector storage, and a FastAPI service to process document images, encode them alongside text, retrieve semantically relevant chunks, and produce accurate, context-aware answers. | |
| Autonomous Research Analyst Agent for Literature Exploration (Ongoing) | Nov 2025 - Present |
| <ul style="list-style-type: none">• Agentic AI research agent to be built on n8n cloud that performs autonomous search-to-report synthesis using modular retrieval and reasoning components. | |
| Multimodal Hybrid xG Model for Football Match Outcome Prediction (Accuyield) | Jan 2025 - Sept 2025 |
| <ul style="list-style-type: none">• Built a multimodal expected goals model using feature engineering, a WD-based input drift detection-correction mechanism and a predictive ensemble comprising of Bayesian-tuned attention-based DL architectures and tree-boosting algorithms, with best-model selection for predicted output for a particular match guided by rolling performance metrics, and outputs adjusted heuristically based on trend and seasonality. This model showed marked improvements in performance over Poisson baselines and enabled analytics that improved ROI by aligning predictions with pre-match odds and match dynamics. | |
| Bank Customer Churn Analysis using ML Methodologies (University of Malaya) | Mar 2023 - Feb 2024 |
| <ul style="list-style-type: none">• Integrated NLP-based SocMed Sentiment Analysis with multi-learning paradigm (TSLs, ELs, DNNLs, DELs) churn modelling, comparing pre- and post-training feature importances to identify robust, model-agnostic churn drivers. XGBoost emerged as the best predictive model for churn across the top 5 Malaysian banks. | |

Cyberbullying Conceptualization, Characterization & Detection in Social Media - A Systematic Literature Review (Sunway-Lancaster University) Jul 2021 - Jul 2022

- Synthesized ML approaches for cyberbullying detection through a SLR of definitions, feature engineering and classification techniques, highlighting model-metric alignment as key to effective automated detection.

Feasibility Study on DL-based Inferencing using Wire Bond Imagery Dec 2021 - Mar 2022
(ACA Vision Technology)

- Assessed DL architectures for wire bond inspection and recommended L-DNN due to organizational constraints around data, infrastructure, and compute, enabling low-data, edge-efficient inference.

Publications

Cyberbullying Conceptualization, Characterization and Detection in Social Media – A Systematic Literature Review Jan 28th, 2023

Woo Wai Hong, Chua Hui Na, Gan May Fen

10.31436/ijpcc.v9i1.374

Professional Experience

Upskilling – Ipoh, Perak, MY Sept 2025 - Present

- Currently undertaking structured online learning to strengthen technical expertise on various topics in AI.

Data Scientist - R&D, Accuyield – Petaling Jaya, Selangor, MY Nov 2024 – Sept 2025

- Co-led applied AI/ML research within an Agile R&D environment through structured experimentation in 3 key areas: Regularization, Architectural Finetuning and Data-centric Optimization - aimed towards improving model performance and mitigating bias.
- Designed a novel bias injection framework leveraging dynamic league standings, alongside a robust input drift detection-correction mechanism which significantly improved model generalizability across varying team-side matchups, with stark improvements in more than 20% of underperforming teams.
- Developed and deployed a multimodal hybrid expected goals (xG) model combining attention-based DL architectures, tree-based boosting, dynamic best-model selection and heuristic-based output adjustment which achieved a 40% reduction in error (vs. baseline Poisson and Ridge Poisson variants).
- Drove data-informed football analytics that contributed to a 7% uplift in ROI by aligning model outputs with pre-match odds and match dynamics.

Jr. Data Scientist/Engineer, Trontal Group – Kuala Lumpur, W.P Kuala Lumpur, MY May 2024 – May 2024

- Supported the Data Engineering team in developing and maintaining dbt models for views creation in ClickHouse for multiple analytics use cases in centralized cryptocurrency exchanges (CEX).
- Developed optimized materialized views for CEX token flow tracking, reducing query runtime by 30% and storage usage by 20%, improving system efficiency and accelerating delivery for front-end and Quant teams to support timely and informed trade executions.
- Co-designed and implemented parameterized views to monitor CEX flows by wallet and by entity over time, enhancing analytical capabilities for identifying token movements tied to known addresses and major trading entities.

Data Science Intern, ACA Vision Technology – Bkt Mertajam, Penang, MY Dec 2021 – Mar 2022

- Conducted foundational research on object detection and instance segmentation for wire-bond defect identification in post-AOI (Automated Optical Inspection) imagery.
- Evaluated and benchmarked deep learning models (Mask R-CNN, Faster R-CNN, Detectron2, YOLOv3) to propose solutions for reducing over- and under-rejection rates within the AOI inspection pipeline.
- Assisted with manual instance mask annotation using WB1, a proprietary wire bond detection platform, to enhance its object boundary detection logic, contributing to a record 95% system-wide accuracy in production.
- Authored a comprehensive feasibility study outlining infrastructure requirements for scalable DL-based inference and delivered a roadmap enabling the organization's transition toward data-centric operations.

Skills

Tech Stack: **Python** (Pandas, NumPy, Matplotlib, Seaborn, Dash, Scikit-Learn, TPOT, Optuna, Streamlit, OpenCV, SciPy, PIL), **R** (caret, randomForest, xgboost, e1071, mlr3, dplyr, ggplot2, tidyverse, shiny, rmarkdown), **Frameworks** (PyTorch, TensorFlow, Keras, HuggingFace), **MLflow**, **SQL**, **Git**, **GitLab**, **Docker**, **Kubernetes (K8s)**, **Tableau**, **Airflow**, **dbt**, **LaTeX**

Hard Skills: Experimental Design, Multimodal and Multi-model AI Systems, Machine Learning, Deep Learning, Model Optimization, CI/CD, Statistical Data Analysis, MLOps, Natural Language Processing, Digital Image Processing, Computer Vision, Data Wrangling, Data Visualization, Prompt Engineering, Vibe Coding, Poster Design

Soft Skills: Creativity, Problem-Solving, Analytical Thinking, Teamwork, Passionate & Quick Learner, Clear Communication, Agile Project Management, Presentation Skills

Personality: Self-assured, Logical, Driven, Positive, Unbiased

Research Interests

Generative Models, Large Language Models (LLMs), Post-Training of LLMs, Design & Optimization of Multimodal and Multi-model AI Systems, Computer Vision, Edge AI, Digital Twins and Virtualization, Trustworthy AI, Human-AI Interaction and Applications in Healthcare, Mobility, Sustainability & in Creative Industries as well as World Models and Artificial General Intelligence (AGI)

Awards

Tan Sri Dato' Seri Jeffrey Cheah Scholastic Award 2023 Feb 2023
Sunway University

Cambridge GCE Advanced Level (A-Level) High Achievers Award 2017 Jan 2018
Sunway College

Extracurricular Activities

Sunway Analytics Society, Member Mar 2019 - Jul 2022
Sunway University

Sunway Tech Club, Member Mar 2019 - Jul 2022
Sunway University

Sunway Leadership, Education and Advancement Programme (LEAP), Member Mar 2019 - Jul 2022
Sunway University

Ipoh Toastmasters Club, Member Jan 2016 - Jun 2016
Toastmasters International

Languages

English: Fluent **Malay:** Proficient **Mandarin:** Elementary **Cantonese:** Elementary

References

Dr. Chua Hui Na, Professor and Programme Leader
Department of Data Science and Artificial Intelligence
School of Computing and Artificial Intelligence
Faculty of Engineering and Technology
Sunway University, Petaling Jaya, MY
huinac@sunway.edu.my

Benjamin Chang Xuan Shi, Lead Data Scientist
Accuyield Sdn. Bhd., Petaling Jaya, MY
benjamin@accuyield.net

Dr. Richard Wong Teck Ken, Senior Lecturer
School of Engineering
Faculty of Engineering and Technology
Sunway University, Petaling Jaya, MY
richardwtk@sunway.edu.my