

Min-Hsien Weng

Email: samminweng@gmail.com

ORCID: <https://orcid.org/0000-0003-2334-1863>

LINKEDIN PROFILE: <https://www.linkedin.com/in/min-hsien-weng-3b753821/>

Areas of specialization

Software Verification; Programming Analysis; Compiler; Parallel Programming

Natural Language Processing; Large Language Model; Data Analytic and Visualization

Skills

Programming: Java / C / Python / C# / JavaScript / PHP / HTML / CSS / OpenCL / Java Threading / Hadoop / Map Reduce / OpenMP / Shell Script / Rust / Dafny

AI: Deep learning (Large Language Model) / PyTorch / TensorFlow / Pandas / NLTK / Stanford CoreNLP / HDBSCAN / BERT (Word2Vec) / Matplotlib / D3.js / Google Chart

IoT: Arduino Uno / ESP32 Board / Sensors / MQTT / IFTTT / RFID / Kiosk

Architect: Compiler / C Memory Management / RFID Controller / Product Data Management

System: Linux / ANT Build / XAMPP / Android / Git / AWS Cloud / Docker

DB: SQLite / MySQL / Oracle SQL Server / MangoDB

Academics: Boogie / SAT Solvers / Theorem Proving / Latex / Haskell / Whiley / Lean

Appointments held

2022-2023	Teaching Fellow, Waikato University Joint Institute at Zhejiang University City College (NZUWI), Hamilton, New Zealand
2019-2022	Postdoc, Waikato University, Hamilton, New Zealand
2014-2019	Doctoral Assistant, Waikato University, Hamilton, New Zealand
2002-2011	Associate Researcher, Industrial Technology Research Institute, Taiwan

Education

2014-2019	PHD in CS, Waikato University, New Zealand
2012-2013	MSc in CS, Waikato University, New Zealand
2011-2012	PGDIP in CS, Waikato University, New Zealand
1999-2001	MASTER in IEEM, National Tsing Hua University, Taiwan
1995-1999	BACHELOR in IEM, National Yang Ming Chiao Tung University, Taiwan

Membership

2023	Full Membership in Sigma Xi, the Scientific Honor Society
2022	Emerging Professional Member of Engineering NZ (Membership number: 2009229)
2023	Reviewer for the open access journal Urban Planning (ISSN: 2183-7635)

Publications

JOURNAL ARTICLES AND CONFERENCE PAPERS

- 2022 **MH Weng**, S Wu, M Dyer, "Identification and Visualization of Key Topics in Scientific Publications with Transformer-Based Language Models and Document Clustering Methods", *Applied Sciences* 12 (21), 11220
- 2021 **MH Weng**, R Malik, M Utting, "Automatic proofs of memory deallocation for a Whyley-to-C Compiler", *Formal Methods in System Design* 57 (3), 429-472
- 2021 **MH Weng**, S Wu, M Dyer, "AI Augmented Approach to Identify Shared Ideas from Large Format Public Consultation", *Sustainability* 13 (16), 9310
- 2021 M Dyer, S Wu, **MH Weng**, "Convergence of public participation, participatory design and NLP to co-develop circular economy", *Circular Economy and Sustainability* 1 (3), 917-934
- 2020 M Dyer, **MH Weng**, S Wu, T Garcia Ferrari, R Dyer, "Urban narrative: Computational linguistic interpretation of large format public participation for urban infrastructure", *Cogitatio* 5 (4), 20-32
- 2019 M Dyer, R Dyer, **MH Weng**, S Wu, T Grey, R Gleeson, TG Ferrari, "Framework for soft and hard city infrastructures", *Proceedings of the Institution of Civil Engineers-Urban Design and Planning* 172 (6), 219-227
- 2019 M Dyer, R Dyer, **MH Weng**, S Wu, T Grey, R Gleeson, TG Ferrari, "Urban narratives for city infrastructures", *WEC2019: World Engineers Convention 2019*, 1127
- 2019 K Mackness, M Dyer, R Dyer, A Hinze, T Garcia Ferrari, S Wu, R Wilkins, **MH Weng**, Urban narrative: Value based urban design", *2019 New Zealand Planning Institute Conference*, 1-24
- 2017 **MH Weng**, B Pfahringer, M Utting, "Static techniques for reducing memory usage in the C implementation of Whyley programs", *Proceedings of the Australasian Computer Science Week Multiconference*, 1-8
- 2016 **MH Weng**, M Utting, B Pfahringer, "Bound analysis for Whyley programs", *Electronic Notes in Theoretical Computer Science* 320, 53-67
- 2013 M Utting, **MH Weng**, JG Cleary, "The JStar language philosophy", *Parallel Computing* 40 (2), 35-50

THESIS AND TECHNICAL REPORT

- 2019 **MH Weng**, "Efficient compilation of a verification-friendly programming language", *Thesis, Doctor of Philosophy (PhD)*, The University of Waikato
- 2019 Dyer, M., Dyer, R., Ferrari, T., **MH Weng**, Wilson, J., Wilkins, R., & Wu, S. "Data Collection, Data Analytics, Data Visualisations and Data Storytelling", *Report for Building Better Homes, Towns and Cities: Urban narratives (Contestable Research)*, 65pgs.
- 2013 **MH Weng**, "Automatic Parallelization of Data-Driven JStar Programs", *Thesis, Master of Science (MSc)*, The University of Waikato

PATTERNS

- 2012 Probability Time Division Multiplexing Polling Method and Wireless Identification Reader Controller Thereof, US US8233468 B2
- 2011 Ubiquitous Proxy Mobile Service Method and System And Computer Recordable Storage Medium For the Method, US US8037130 B2

CERTIFICATES

2023	Machine Learning Specialization, Coursera
2022	Fundamentals of Deep Learning, NVIDIA
2010	Cloudera Certified Hadoop Developer (CCHD), Cloudera

Teaching

2022-23	Teaching Fellow at Waikato University Joint Institute at Zhejiang University City College (NZUWI) Taught the following Computer Science courses: COMPX201 Data structures and algorithms), COMPX222 (Web development), and COMPX318 (Internet of Things and Mobile computing).
2020	Senior Tutor (6 months) at NZUWI Taught COMPX322 (Advanced Web Development) and COMPX202 (Mobile Computing)
2014-2019	Doctoral assistant at Waikato university helps out a variety of undergraduate and postgraduate courses.

Research Highlights

Kaggle Notebook and Discussion Expert (top 500 of 326,000 participants from Sep 2023)

LLM - Detect AI Generated Text fine-tuning large language models (LLMs) on 1M+ machine-generated texts to detect AI content. My solution received hundreds of up-votes and forks, and led me to the Kaggle Expert recognition. [\[Gold-medal Solution\]](#).

Google - predict AI Model runtime developing BERT and graph neural networks (GNNs) for AI model runtime prediction, hosted by Google. [\[Bronze-medal solution\]](#)

LLM Science Exam - uses LLMs to answer difficult science questions blends Retrieval Augmented Generation (RAG) with large language model Platypus2-70B. RAG retrieves relevant Wikipedia texts, which are then utilized by Platypus2-70B to generate highly accurate answers to diverse questions. [\[Solution link\]](#)

A large language model-based (GPT-3) topic analysis tool [Postdoc Project](#)

AI augmented approach to identify ideas from large public consultation [Postdoc Project](#)

A compiler translates into efficient C code and formalize and prove the safety of C code [PhD Project](#)

Pavilion of Dreams at 2010 Flora Expo (the International Association of Horticultural Producers). I developed the software architecture that seamlessly integrates our custom RFID hardware, empowering visitors to unlock diverse interactions with exhibits in the building using their RFID bracelets. The exhibition drew over 100,000 visitors in just six months and it is also the first internationally recognized exposition that utilizes RFID technology in Taiwan and the seventh in Asia. Notably, two patents were filed during the project.