Yiming /i' miŋ/ PENG

Data Engineer/Data Architect/Software Engineer

Cellphone: (+64)220619380 | Email: yimingpengjojo@gmail.com | Github: yimingpeng | Linkedin: yiming-nz

OVERVIEW

- **8+ years** of hands-on expertise with all parts of the data integration life cycle, from requirement gathering through data modelling, ETL implementation, reporting, and dashboarding, as well as maintenance and support. As a lead data engineer and principal data architect, I've worked on a variety of projects, including developing ETL pipelines, migrating data warehouses to the cloud (specifically AWS), enterprise systems, and constructing robust DataOps/MLOps platforms.
- **20+ years** programming experience with expertise in:
  - Clean, efficient code development in Python, Java, MATLAB, and SQL across 60+ industrial and academic projects
  - Large-scale system debugging and development using frameworks including TensorFlow, PyTorch, Keras, JADE, and scikit-learn
- **18+ years research and hands-on experience on designing, developing, and applying various Artificial Intelligence and Machine Learning programs/tools for solving various problems, such as**
  - Computer Vision and Pattern Recognition: image classification, network behavior anomaly detection, and financial market forecast.
  - Reinforcement Learning: robotic control and intelligent game-play.

  A full list of 15 publications is available on request.
- **16+ years** working experience in New Zealand and **5 years** postgraduate studying experience in Computer Science of leading NZ universities providing me the ability to produce high-quality written contents for both technical reports and scientific papers.
- **8+ years** teaching and tutoring experience, extensive public speaking experience, and considerable collaboration experience with international industrial partners (e.g., National Institute of Communication and Technology, NICT Japan) enabling comfortable communication with different audiences such as non-technical end-users, technical professionals, scientific researchers, and business stakeholders.

TECHNICAL SKILLS

Programming	Python, SQL, Java, MATLAB, R, HTML, JavaScript, CSS, C, C++
Data Analytical Tools	Pandas, Numpy, Scipy, Matplotlib, Seaborn, Bokeh
Data Engineer Tools	Airflow, DBT, Dagster, N8N
Business Intelligence Tools	Tableau, PowerBI, Qlik
Machine Learning Tools	Scikit Learn, TensorFlow, Keras, Theano, PyTorch, Deeplearning4J
Big Data Tools	Spark, Hadoop, Hive
Cloud Computing Platform	Amazon Web Services (AWS), Microsoft Azure, Google Could Platform (GCP), NeSI High Performance Computing (HPC)
Databases/Data Warehouses	Snowflake, Redshift, PostgreSQL, Oracle, SAP Hana, MySQL, Microsoft SQL Server, Salesforce
Text Editing	Latex, Markdown, Microsoft Office
Other Tools	Git, Docker, Confluence, K8S, OpenShift, Azure Pipelines, BitBucket, Gitlab

WORKING EXPERIENCE

Unity & Wētā Digital (Wētā FX) Wellington, New Zealand

Senior Software Engineer - Data Oct. 2023 – Present

- Projects:
    - Developing and managing an on-premises DataOps platform that includes 25+ ETL pipelines, focusing on improving data ingestion/transformation and reporting in Tableau. Tech stacks: Python, SQL, Ansible, Gitlab Pipelines, MySQL, PostgresSQL, Clickhouse, Tableau.
    - Led design and development of a data quality monitoring system using Great Expectations, improving observability across the entire data platform. Tech Stacks: Python, Great Expectations, SQL, Ansible, Gitlab Pipelines, PostgresSQL, Clickhouse.
    - Building a resource planning tool for managing and scheduling crew workloads and capacity optimization. Tech Stacks: Python, ReactJS, SQL, Ansible, Gitlab Pipelines, PostgresSQL.
    - Created real-time dashboards and analytics reports for Talent & Culture department to track Return to Office (RTO) policy implementation and organizational compliance. Tech Stacks: Python, SQL, PostgresSQL, Tableau.
    - Contributing to productionizing machine learning models for predicting workflow schedules and budgets in production environments. Tech Stacks: Python, SQL, PostgresSQL.
    - Collaborating with data practitioners to form enterprise BI data strategy for the organization.
    - Developed and maintained multiple production-grade applications for internal stakeholders. Tech Stack: Python, SQL, ReactJS, PostgresSQL, MongoDB.
- Deliverables:
    - 6 production-ready ETL pipelines
    - 1 Data Quality Check/Monitor platform
    - 1 strategic RTO implementation analysis report

Emergence Caller Location Information at MBIE Wellington, New Zealand

Data Architect Jul. 2022 – Jul. 2023

- Data Analytical Projects:
    - Successfully migrated, redesigned, and maintained 8 data ETL pipelines from a Docker-compose solution to OpenShift, utilizing various technologies such as Airflow, DBT (dbt-expectations, dbt-elementary), Dynatrace, OpenShift, and Azure Pipelines.
        - Virtual Mobile Location Center (vMLC REST API) to PostgresSQL
            - Base Station Almanacs
            - Location Detail Records
        - GSMA Device Database (REST API) to PostgresSQL
        - Genuine Call Records (SFTP) to PostgresSQL
        - Emergency Call Records (Flat Files) to Postgres SQL
            - Fire Emergency
            - St Johns
            - Wellington Ambulance
            - NZ Police
    - Designed a modern DataOps platform, encompassing Data Observability, Data Catalog, Data Linage, Data Management, Data Governance and Data Strategy.

- Deliverables:
    - 8 data ingestion pipelines
    - 1 comprehensive design of a modern DataOps Platform

IHC New Zealand Wellington, New Zealand

Principal Data and Integration Architect Jul. 2021 – Jul. 2022

- Data Analytical Projects:
    - Designed, developed, and maintained an automated user provisioning process, from ServiceNow/Payglobal to Azure Active Directory. Techniques include Python (Airflow), Snowflake, REST API.
    - Designed and implemented an AWS Elastic Container Services based data platform for data analytical tools including Retool, N8N, and Airflow. Techniques include AWS CloudFormation, Python, and Docker.
    - Designed, developed, and maintained multiple ETL pipelines by using Airflow, DBT (dbt-expectations, dbt-elementary)
        - Kronos (MySQL) to Snowflake
        - MySupport (MySQL) to Snowflake
        - Payglobal (MSSQL) to Snowflake
        - OneChart (Flatfiles in SFTP) to Snowflake
        - GreenTree/RiskManager/Business Central/Azure AD (REST API) to Snowflake
    - Built 33 Power BI reports/dashboards for different analytical purposes. Techniques include Power BI, Snowflake.
- Deliverables:
    - 9 data ETL pipelines
    - 33 Power BI reports/Dashboards
    - 1 generic DataOps Platform.

Chorus New Zealand Ltd Wellington, New Zealand

Senior Data Engineer Dec. 2019 – Jun.2021

- Data Analytical Projects:
    - Designed, developed, and maintained multiple internal data ingestion pipelines using Python (Apache Airflow)/PySpark:
        - Oracle to Redshift/Snowflake
        - SAP Hana to Redshift/Snowflake
        - Flat Files to Redshift/Snowflake
        - JSON/XML to Redshift/Snowflake.
    - Designed, developed and maintained an internal financial data analytical system. Techniques include Python, SAP Hana, and Tableau.
    - Designed and developed an internal data warehousing replication system from Oracle Data Warehouse to AWS Redshift using Python, Spark.
    - Worked with other colleagues to design and build a Kubernetes based MLOps platform, techniques include CloudFormation, AWS EKS, Kubeflow.
- Deliverables:
    - 5 data integration pipelines.
    - 1 financial data analytical system

- 1 generic MLOps Platform

KPMG New Zealand Wellington, New Zealand

Data Engineer Mar. 2019 – Nov.2019

- Data Analytical Projects:
  - Co-developed/Maintained three data analytical projects for three different government agencies and four different companies, including suspicious transaction analysis, financial data modeling, financial transaction continuous monitoring, and geo-location analysis. Techniques include Python, R, SQL and NodeJS.
  - Co-developed/Maintained two cloud-based data analytical platforms (AWS and Azure) for two different government agencies. Among them, the platform built on Azure is particularly high-profile national-wide project. Techniques include NodeJS and SQL.
- Deliverables:
  - 2 data analytical platforms
  - 1 data enrichment system with web scraping

Victoria University of Wellington Wellington, New Zealand

Research Scientist and Research Assistant Sep. 2015 – Mar. 2019

- Data Analytical/Machine Learning Projects:
  - Implemented/Co-implemented a number of new Reinforcement Learning systems (more than 100,000 lines of code) using Python, Theano, TensorFlow, PyTorch, Numpy, Scipy, Scikit-Learn, Pandas, Java, DeepLearning4J, and MATLAB.
  - Implemented a new deep learning based ensemble system for image classification (~3,000 lines of code) using Python, TensorFlow, Keras, Numpy, Pandas and Scikit-Learn.
- Conference Technical Support:
  - Collaborated with colleagues for organizing two international conferences, i.e., the 31st Australasian Joint Conference on Artificial Intelligence (AI 2018) and 2019 IEEE Congress on Evolutionary Computation (CEC 2019).
  - Developed and maintaining the two conference websites (more than 20,000 lines of code) using Bootstrap 3 with HTML5, CSS3.0, and JavaScript.
  - Managing 4 social media channels of CEC 2019 (i.e., Facebook, Twitter, Linkedin, and Google+) for publicity of the conference.
- Teaching Technical Support:
  - Designed/Developed a CNN based Image Classification framework as a course project template (~1,000 lines of code) using Python, Keras, Scrapy, and Scikit-Learn.
  - Designed/Developed an Android mobile image retriever APP as a course project template (~1,500 lines of code).
  - Designed/Developed a J2EE WebService application as a course project template (~1,500 lines of code) using J2EE, Tomcat 6.0.
- Deliverables:
  - 15 reinforcement learning systems (Github Link).
  - 1 CNN based image classification system.
  - 3 course projects.

- 2 conference websites, i.e., AI 2018 and CEC 2019.

Unitec Institute of Technology Auckland, New Zealand

Research and Teaching Assistant Sep. 2011 – Jun. 2015

- Data Analytical/Machine Learning Projects:
  - Led/Coordinated a technical team (2 Ph.D. students and 2 master students) to develop a decentralized online network traffic monitoring system with real-time visualization (more than 150,000 lines of code) using J2SE, Java Agent Development Framework (JADE), Unity3D and MySQL, which remains in use at Unitec till today.
  - Implemented/Co-implemented the financial forecast systems for stock and crude oil market trend prediction (~1,000 lines of code) using MATLAB.
- System Technical Support:
  - Deployed three real-time network traffic monitoring systems ``NICTER'', ``NIRVANA'' and ``NIRVANA Rev. 2'', in collaboration with three colleagues from National Institute of Communication and Technology (NICT), Japan.
  - Maintained/ensured the smooth operation of the above three systems from 2012 to 2015 in the Cyber Security Research Center at Unitec, which later in 2015 contributed as one of the core projects for a $10.6 million (excl. GST) cyber-security research project funded by Ministry of Business, Innovation, and Employment (MBIE) and led by the University of Waikato.
- Teaching Technical Support:
  - Developed 6 Android mobile APPs for an outreach teaching project.
- Deliverables:
  - 1 decentralized online network traffic monitoring system.
  - 2 financial forecast systems.
  - 6 Android mobile APPs.

PROFESSIONAL CERTIFICATES

- CKAD: Certified Kubernetes Application Developer | The Linux Foundation | Jun 2025 - Jun 2027
- Tableau Server Certified Associate | Tableau | May 2025 - May 2027
- Astronomer Certification DAG Authoring for Apache Airflow | Astronomer | Oct 2022
- Astronomer Certification for Apache Airflow Fundamentals | Astronomer | Aug 2022

QUALIFICATIONS

Victoria University of Wellington (VUW) Wellington, New Zealand

Ph.D. in Artificial Intelligence Sep. 2015 – June. 2019

- *Research Directions*:
  - Artificial Intelligence and Machine Learning (specifically Reinforcement Learning, Deep Learning, Feature Learning, Transfer Learning, Ensemble Learning, Image/Vision Recognition)
  - Numerical Optimization and Convex Optimization (specifically Gradient Descent, Heuristic Search, Evolutionary Computation)
  - Big Data and Data Analytics (specifically statistical analysis, data acquisition, data manipulation, data wrangling, data cleansing, data visualization)
- *Thesis*: Policy Direct Search for Effective Reinforcement Learning

- *Advisors*: Dr Aaron Chen and Prof. Mengjie Zhang

Auckland University of Technology (AUT) Auckland, New Zealand

Master of Computer and Information Sciences Jun.2009 – Jul.2011

- *Research Directions*: Artificial Intelligence and Machine Learning (specifically Supervised Learning, Feature Extraction/Construction, Image/Vision Recognition)
- *Thesis*: Boosting performance of incremental IDR/QR LDA - from sequential to chunk
- *Advisors*: Prof. Alvis Fong and Prof. Shaoning Pang
- *Grade*: A+ (First Class First Division), ranked as the Top 1 Master of Computer and Information Sciences in 2011

Wuhan University of Technology Hubei, China

Bachelor of Computer Science and Technology Sep.2000 – Jul.2004

- Key Courses: C Programming Language, Data Structures and Algorithms (ANSI C), Object Oriented Programming (C++), Database Design (SQL), Software Engineering, Requirement Engineering, Operating System Design, Computer Networks, Assembly Language (IBM 8086), Artificial Intelligence.
- GPA: A (grade)

AWARDS & SCHOLARSHIPS

- 2019 KPMG New Zealand People's Choice Award
- 2018 ACM SigEVO Student Travel Award ($700 NZD)
- 2018 Victoria Thesis Submission Scholarship, VUW ($6,000 NZD)
- 2018 Faculty Strategic Research Grant, VUW ($3,000 NZD)
- 2017 The 26th International Joint Conferences on AI (IJCAI 2017) Student Travel Grant ($1,000 USD)
- 2017 Faculty Strategic Research Grant, VUW ($3,000 NZD)
- 2015 Victoria Doctoral Scholarship, VUW ($23,500 NZD p.a. for three years plus tuition fee)
- 2014 Nominee for Highest Rated Lecturer, Unitec
- 2011 Eagle Technology Scholarship ($2,500 NZD for the top Master graduate)

RESEARCH EXPERIENCE

Victoria University of Wellington Wellington, New Zealand

Research Scientist and Research Assistant Sep. 2015 – Present

- Investigated/studied various Artificial Intelligence/Machine Learning techniques involving Optimization, Deep Learning, Reinforcement Learning, Evolutionary Computation and different Search Mechanisms to design/develop 15 different Machine Learning systems (more than 10,000 lines of code for each system) that can
    - automatically extract/construct high-level features from raw robot sensor inputs or game screen-shot images,
    - efficiently and effectively maneuverer simulated robot controllers/game controllers to accomplish difficult tasks (e.g., controlling 2-leg robot to walk, playing Atari games, or controlling temperature in an air-conditioning system), or
    - be theoretically/mathematically proven to converge, and

    - all the new systems I developed perform better than existing systems (e.g., the systems developed by either OpenAI or Google).
- Investigated/studied various Machine Learning/Data Analytical techniques involving Deep Learning, Feature Learning, Transfer Learning, Ensemble Learning, Classification, Dimension Reduction, Data Wrangling, Data Cleansing, Data Augmentation and other Data Analytical methods to design/develop a deep Convolution Neural Networks based image classification system that can
    - accurately and automatically detecting false alarms from low-quality raw images collected from remote cameras deployed in many bushes throughout NZ,
    - significantly improve the classification accuracy from ~71% (InceptionV3 developed by Google) to ~84% (our system) with a limited amount of training data, and
    - largely save a huge amount of time for ecologists who used to manually check such false alarms.
- Deliverables:
    - 10 conference publications.
    - 2 new conference manuscripts in preparation.
    - 2 journal manuscripts submitted for review.
    - 1 new journal manuscript in preparation.

Unitec Institute of Technology Auckland, New Zealand

Research and Teaching Assistant Sep. 2011 – Jun. 2015

- **Real-time Network Traffic Analysis**: Investigated/studied various Machine Learning/Data Analytical techniques involving Clustering, Dimension Reduction, Classification, Decentralized Computing, Data Acquisition, Data Wrangling, Data Cleansing, Data Visualization, and other Data Analytical methods to design and develop a decentralized online network traffic analysis system with real-time visualization that can
    - automatically detect anomaly packets from online network traffic,
    - achieve 90% recognition rate within a simulated network environment, and
    - perform approximately three times faster than the centralized baseline system.
- **Financial Forecast**: Worked collaboratively with a group of 3 researchers to investigate/study various Machine Learning/Data Analytical techniques involving Correlation Analysis, Regression Analysis, Prediction, Model Selection, Data Preprocessing, Data Mining, and other Data Analysis methods to develop financial (stock market and crude oil market) data analysis systems that can
    - automatically acquire public financial data from the Internet,
    - efficiently identify the correlations among the data, and
    - better predict the financial trends in comparison to some existing systems.
- Deliverables: 2 conference publications.

Auckland University of Technology Auckland, New Zealand

Part-time Research Assistant Sep.2010 – Jul.2011

- Investigated/studied various Machine Learning/Data Analytical techniques involving Image Classification, Feature Extraction/Construction, Image Pre-processing, Statistical Learning, and other Data Analysis methods to a new Linear Discriminant Analysis (LDA) feature extraction based face recognition system that can
    - significantly improve the classification rate from ~85% to ~95% on some benchmark datasets, and
    - performs approximately two times faster than most existing systems on six benchmark datasets.
- Deliverables: 1 conference publication.

REFERENCES

Mischa Douglas	**Kieran Hume**
Senior Project Manager	Data Analytics Manager
Victoria Police	Chorus Ltd
Australia	New Zealand
✉ mischa.douglas@police.vic.gov.au	✉ Kieran.hume@chorus.co.nz
☎ +64273574466	☎ +64 21306396

Luis Ubaldo	**Phil Parker**
Principal Data Engineer	Director at ECLI
IHC	MBIE
New Zealand	New Zealand
✉ luis.ubaldo@ihc.org.nz	✉ phil.parker@mbie.govt.nz
☎ +64 226217327	☎ +64 275316750