

# Mia Yang

Glenfield, Auckland • +64 27 287 4925 • miayangyyx@gmail.com • linkedin.com/in/yuxia-mia-yang

---

Graduate Data Engineer with strong foundations in machine learning, NLP, and GenAI system design. At One NZ, I contributed to projects across Streamlit, Power BI, and Snowflake, helping teams improve efficiency and drive data-informed decisions. I have built LangChain agents, RAG-based document assistants, and integrated GPT APIs into internal applications. Passionate about building scalable, responsible AI workflows that solve real problems.

## KEY SKILLS

---

### Machine Learning & AI: TensorFlow, Scikit-learn, PySpark

Solid academic foundation with formal training in supervised/unsupervised learning, ensemble methods (random forest, boosting), and deep learning using TensorFlow/Keras. Applied ML in both academic and real-world projects including Fashion-MNIST image classification, load forecasting, and NLP tasks.

### Generative AI & Agents: LangChain, APIs, n8n, CrewAI

Built RAG chatbot and CSV agent with contextual memory and response filtering. Integrated GPT APIs into internal applications and web tools.

### Software Development & Integration: Python, SQL, R, Git

Familiar with Java; able to ramp up quickly as needed. Proficient in prompt engineering, with hands-on practice through Vibe Coding and personal apps that have been deployed and maintained.

### Data Engineering & Platforms: Snowflake on AWS, dbt, Streamlit, Power BI, Azure DevOps (CI/CD)

Experience building and deploying Snowflake-native apps, real-time dashboards, and cost attribution models. Familiar with Snowflake on AWS; currently expanding AWS skills (AI Practitioner certification in progress).

### Soft Skills & Cultural Fit

Known for collaborative problem-solving, curiosity in AI innovations, and strong communication with technical and non-technical stakeholders. Regularly contribute to Agile ceremonies and technical documentation. Actively follow responsible AI practices, such as hallucination reduction, memory control, and data transparency. Passionate about bridging the gap between cutting-edge AI research and real-world application, aligned with One NZ's "people serving people" value

## WORK EXPERIENCE

---

### One New Zealand

Auckland, New Zealand

Data Engineer (*Intern* → *Part-Time*)

11/2024 - Present

- Led end-to-end development of a Snowflake-native Streamlit app for handset data workflows, enabling real-time updates and reducing manual reporting time by 40%
- Delivered real-time dashboards with Power BI + Snowflake for Anodot Replacement project, enhancing cross-team visibility
- Designed a Snowflake cost attribution model that connected platform usage with business teams, improving cost transparency and enabling data-driven FinOps decisions.
- Collaborated across Network Squad, Commercial, and Data Engineering; contributed to Agile planning and documentation

### Yes Energy, New Zealand

Data Scientist Intern

Jun.2024 – Jul.2024

- Applied time-series models for load forecasting and presented findings to analysts.

## PROJECTS

---

### GPT-Powered RAG Chatbot

- Built a chatbot using OpenAI, and vector search for internal documentation Q&A.
- Tuned chunking and retrieval logic, reduced hallucinations through response filtering and memory control.

### Snowflake Cost Agent

- Developed a GPT-based agent capable of answering CSV file queries with memory-enabled context retention.
- Added guardrails to handle malformed input and unsupported file formats.

### Climate Data Analysis Using Spark

- Processed 3+ billion rows of global climate data using PySpark, SQL, and Python.
- Cleaned, joined, and analyzed data for visualizing trends across 263 years.
- Created global precipitation map and contributed to environmental analytics.

## EDUCATION

---

### Master of Applied Data Science (Grade: A)

University Of Canterbury

Christchurch, New Zealand • 02/2024 - 06/2025

#### Relevant Coursework in AI & Generative Technologies:

DATA420: Scalable Data Science – Processed big data using PySpark; evaluated ensemble methods such as random forest and boosting on real-world datasets.

DIGI405: Texts, Discourses and Data – Fine-tuned DistilBERT for sentiment analysis; implemented topic modelling (LDA) and clustering (k-means) for NLP applications.

STAT448: Big Data – Built and tuned deep learning models (e.g., CNNs) for image classification using TensorFlow/Keras.

STAT462: Data Mining – Applied supervised and unsupervised learning methods including decision trees, clustering, and ensemble models.

### Bachelor of Accounting

Nanchang University

Nanchang, Jiangxi, China • 09/2016 - 06/2020

## CERTIFICATIONS

---

Snowflake SnowPro Core – Passed Dec 2024

AWS AI Practitioner – In Progress (expected Aug 2025)

## PROFESSIONAL DEVELOPMENT

---

- 🏆 2nd Place – AI Hackathon (AI Forum NZ, South Island)
- 🎤 Toastmasters Member – Communication & Public Speaking
- 📍 Active participant in AI & Data meetups and workshops

**References available upon request.**