

# Mia Yang

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

Graduate Data Engineer with practical experience in machine learning, NLP, and GenAI system design. At One NZ, I delivered 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 applications. Passionate about building scalable, responsible AI workflows that solve real problems.

## KEY SKILLS

### Machine Learning & GenAI:

**TensorFlow, Scikit-learn, LangChain, LangGraph, CrewAI, n8n, APIs**

Practical experience with deep learning, classic ML, and NLP (DistilBERT, LDA, k-means) in academic projects. Developed GenAI agents and RAG chatbots in personal projects to explore multi-agent orchestration and retrieval-augmented workflows.

### Programming & Frameworks:

**Python, SQL, R, PySpark, 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.

### Cloud Platforms & DevOps:

**Snowflake on AWS, dbt, Streamlit, Power BI, Azure DevOps (CI/CD)**

Strong hands-on experience with Snowflake—including app development, dashboard integration, and cost attribution. Familiar with deployment automation and currently studying for AWS AI certification.

### Responsible AI & Collaboration:

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

<b>One New Zealand</b>	<b>Auckland, New Zealand</b>
<b>Data Engineer (<i>Intern</i> → <i>Part-Time</i>)</b>	<b>11/2024 - Present</b>
<ul style="list-style-type: none"><li>• 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%</li><li>• Developed the Power BI + Snowflake dashboards for the Anodot Replacement project, contributing to a team solution that saves ~\$60K per year</li><li>• Designed a Snowflake cost attribution model that connected platform usage with business teams, improving cost transparency and enabling data-driven FinOps decisions.</li><li>• Collaborated across Network Squad, Commercial, and Data Engineering; contributed to Agile planning and documentation</li></ul>	
<b>Yes Energy, New Zealand</b>	
<b>Data Scientist Intern</b>	<b>Jun.2024 – Jul.2024</b>
<ul style="list-style-type: none"><li>• Built and evaluated time-series models for accurate energy demand forecasting</li><li>• Applied feature engineering and automation using Bash scripts to improve data pipeline efficiency</li></ul>	

## PROJECTS

---

### Resume Optimization with CrewAI

Worked on a CrewAI-powered multi-agent system for resume-job matching. Studied agent coordination across Job Analysis, Resume Scoring, and Company Research, and ran the full pipeline using OpenAI and Serper APIs. Gained hands-on understanding of GenAI orchestration, agent role design, and collaborative task flow. Applied the tool on real job postings with my own resume for targeted optimization and interview prep.

### Snowflake Cost Agent (LangChain + Streamlit)

Developed an interactive GPT-based agent using LangChain to analyse and visualise CSV-formatted Snowflake cost data. Enabled natural language queries and interactive visualizations via Streamlit. Focused on structured data reasoning and internal cost analysis support.

### GPT-Powered RAG Chatbot

Built a chatbot that answers questions from internal PDF documentation using vector search and content chunking. Developed retrieval logic to enhance context relevance and applied prompt tuning to reduce hallucinations. Applied OpenAI embeddings and a custom semantic search pipeline.

### Climate Data Analysis Using Spark

Processed over 3 billion rows of global climate data using PySpark, SQL, and Python. Designed scalable data pipelines and applied geospatial visualization to support large-scale environmental analysis.

## 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 tasks.

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 algorithms, model evaluation, and a brief overview of reinforcement learning.

### 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.**