WONG MING JIE

Email: mingjwong@hotmail.com | GitHub: @mingjiewong | Huggingface: @mjwong |
LinkedIn: https://sg.linkedin.com/in/ming-jie-wong | Phone: +65 97369640
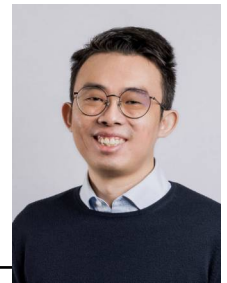
Technical Skills

Programming & Scripting
- Proficient in Python for automation, scripting, and machine learning
- Basic knowledge of C and C++ for systems programming

DevOps Tools
- Experience with Terraform for Infrastructure as Code (IaC)
- Proficient in Kubernetes for container orchestration and Docker for containerization
- Experience with GitHub Actions for automating CI/CD pipelines
- Experience with Argo CD for managing GitOps-based continuous delivery workflows

Data Management & Cloud Platforms
- Experience with ELT pipelines, cloud storage (Azure Blob Storage), workflow orchestration (Kestra), and SQL/NoSQL integration (Postgres, MongoDB) on Kubernetes (AKS)
- Proficient in deploying and managing applications on AWS, Azure and GCP

Experience

Data Scientist,   **Mar 2025 - Jul 2025**
Volvo Cars Tech Hub Singapore
- Provisioned an Azure Kubernetes Service (AKS) cluster and implemented GitOps with Argo CD, which supports production workloads such as the data sanitization pipeline
- Deployed self-hosted LLMs (e.g., Qwen, Gemma, DeepSeek, Mistral etc.) on RTX 6000 GPU cluster to provide internal teams with a private, cost-efficient alternative to commercial LLM APIs (e.g., GPT-4, Gemini etc.)
- Supported the PII-compliant data sanitization pipeline, enabling safer reuse of customer data for internal experimentation and model development
- Prototyped ELT pipeline orchestration with Kestra on AKS cluster, targeting a multi-layered (bronze/silver in Azure Blob Storage, gold in MongoDB) social media data lake for customer feedback analysis

Data Scientist,   **Nov 2021 - Nov 2024**
Singapore Press Holdings (SPH) Media Trust
- Reduced article-tagging API latency by over 66% (from over 30s to under 7s) by developing, deploying and testing REST services in production for multiple publications, including *The Straits Times*, *The Business Times*, *Lianhe Zaobao*, *Tamil Murasu*, and *Harper's Bazaar*
- Reduced article length by 85% (~1,000 to ~150 words) and cutting average reading time from 5 minutes to under 1 minute by designing and deploying end-to-end text summarization pipelines (GPT-4 and Gemini-Pro) for *The Business Times, Berita Harian*, *Harper's Bazaar*, and *Her World*
- Architected the company's first MongoDB Atlas data store, ingesting and enriching 500 K+ articles with metadata (entities, sentiments, semantic vectors, content categories) to support article profiling, semantic search, and personalized recommender systems
- Implemented Knowledge-Graph-driven multi-document summarization with Gemini-Pro, automating generation of monthly e-newsletters and cutting prompt input size by 60-80%
- Co-developed a LightGCN-based recommender for The Business Times, enabling readers to "follow" keywords and receive personalized article suggestions

Graduate Research Assistant,   **Aug 2018 – Mar 2021**
NUS School of Computing

- Applied machine learning and econometrics to tackle societal and business challenges on online social networks during doctoral research in Information Systems (PhD program, incomplete)

Statistician, **Feb 2017 – June 2018**
Department of Statistics, Ministry of Trade and Industry
- Supported Singapore's 11th Household Expenditure Survey 2017/18 for 13,100 households

Open Source Contributions

Open-Source Text Classification Models **May 2023 – Present**
- Developed and open-sourced zero-shot mono- and multilingual classification models on Hugging Face using advanced Large Language Models (LLMs) trained on Natural Language Inference (NLI) datasets for versatile applications
- Achieved over 39,000 all-time downloads (as of July 2025), demonstrating widespread utility and impact
- Customised these models for integration into multilingual applications, enabling language-agnostic systems to support underrepresented languages like Swahili, Urdu, and Thai

AI-Powered Biodiversity Explorer – Ecologist **Feb 2025 – Present**
https://huggingface.co/spaces/mjwong/ecologist
- Designed and developed an AI tool "Ecologist" that identifies wildlife species from photos using multimodal image retrieval and BioCLIP-based visual encoders
- Integrated country-specific biodiversity databases (Singapore, Malaysia, Indonesia and Thailand) to match uploaded images with scientific/common names and visually similar species
- Designed the system for fast species recognition and interactive visual exploration to support biodiversity education and awareness

Singlish-Tuned Automatic Speech Recognition (ASR) Models **Apr 2025 – Present**
https://huggingface.co/spaces/mjwong/Whispeh-Leh
- Fine-tuned OpenAI Whisper and Microsoft Phi-4 models on 66.9k audio-transcript pairs from IMDA's National Speech Corpus
- Adapted models for Singlish speech patterns through selective parameter unfreezing, prompt engineering and supervised fine-tuning
- Achieved a competitive 13.16% Word Error Rate (WER) on SASRBench-v1, a self-designed benchmark for evaluating ASR performance on Singlish, outperforming several baselines
- Deployed and open-sourced models on Hugging Face and Spaces, making them accessible to researchers and developers and achieving over 1,300 all-time downloads (as of July 2025)

Education

Master of Science in Computer Science, **Jan 2024 – Present**
Georgia Institute of Technology (GT)

BSc (Hons) in Mathematical Sciences, **Aug 2012 – Aug 2016**
Nanyang Technological University (NTU)

Certificates

- Certified Kubernetes Administrator (Expires Jan 2026; re-certification in progress)
- Certified Kubernetes Application Developer (Expires Aug 2026)
- AWS Certified Solutions Architect Professional (Expires Aug 2025)

Selected Conference Presentations (during pursuit of PhD in Information Systems)

- Wong M. J., Oh H., Goh K. Y. (2020) "The Dis-United States: Evidences of Polarization in Social Media Speech from Political Elites", Symposium on Statistical Challenges in Electronic Commerce Research, Virtual Conference, June 18-20