

# Shih-Yi (Matteo), Soo

Homepage: [matteosoo.github.io](https://matteosoo.github.io) ◇GitHub: [github.com/matteosoo](https://github.com/matteosoo)  
Phone: (+886) 981-359-060 ◇Email: [matteosoo11@gmail.com](mailto:matteosoo11@gmail.com)

## SUMMARY

Senior Machine Learning Engineer with 4+ years of experience in Speech Processing and NLP. Specialized in building production-grade GenAI/RAG systems on GCP. Led Taiwan's first financial conversational AI from PoC to launch (July 2025), serving millions of users with robust safety architectures and scalable hybrid-cloud deployment.

## WORK EXPERIENCE

- E.Sun Financial Holding Co., Ltd**  
*Senior Machine Learning Engineer*

Taipei, Taiwan  
*Mar 2023 - Present*
- **GenAI Chatbot (i-chat):** Led the E2E development and launch of Taiwan's first financial GenAI service; designed a multi-layer RAG safety architecture with Google Model Armor and policy guards to ensure zero-hallucination.
  - **Hybrid RAG Engine:** Built a retrieval system merging internal research with web search + reranker; optimized response routing via parametric similarity gap modeling and migrated architecture to GCP (Cloud Run, AlloyDB).
  - **Technical Leadership:** Managed a 4-person core team; coordinated across app/data units; demoed solutions at **NVIDIA GTC 2025** and **FinTech Taipei**, featured in 50+ media outlets.
  - **AI Community:** Organized one of Taiwan's largest RAG Hackathons (487 teams); innovated automated evaluation using GPT-4, establishing an enterprise benchmark for GenAI quality assessment.

- E.Sun Financial Holding Co., Ltd**  
*Machine Learning Engineer*

Taipei, Taiwan  
*Jul 2021 - Mar 2023*
- **Speech (TTS):** Built in-house TTS service (RTF 0.3) using Tacotron2/VITS with 20-hour custom dataset; fine-tuned IPA-based models to support 4-language interleaving for telephony IVR systems.
  - **NLP & Infrastructure:** Managed 12+ BERT-based chatbots on Kubernetes; deployed high-availability gateway APIs handling 1000+ monthly calls, significantly reducing maintenance overhead.
  - **RecSys:** Developed a mobile banner recommender for 9M users using DeepFM/Spark (AUC: 0.594, Recall: 0.920).

- Walsin Lihwa Corporation**  
*Software Engineer - Internship*

Taipei, Taiwan  
*Apr 2018 - Jul 2018*
- Applied PageRank and bi-LSTM for industrial text summarization; automated HR web scraping for talent acquisition.

## TECHNICAL SKILLS

ML & AI	RAG, LLM, Agent-based Modeling, DeepFM, PyTorch, TensorFlow, BERT
Speech & NLP	Text-to-Speech (TTS), IPA Encoding, Knowledge Engineering, Text Summarization
Cloud & Infra	Google Cloud (GCP/Cloud Run/AlloyDB), Kubernetes, Docker, Spark, Airflow
Dev Tools	Python, C++, Java, FastAPI, Git, SQL, Linux

## EDUCATION

National Tsing Hua University	2018 - 2021
M.S. in Computer Science	Hsinchu, Taiwan
Soochow University	2014 - 2018
B.S. in Computer Science and Information Management	Taipei, Taiwan