

Bo-Han (Sam) Huang

 SamHuang1018 |  sam-huang |  ju810609@email.com |  +886-930-134-819

SUMMARY

AI / Machine Learning Engineer specializing in LLM training and deployment, speech recognition, and generative AI applications. Experienced in building on-premise AI solutions for financial industry, including intelligent meeting assistants, customer service bots, and document recognition systems. Proven track record deploying production ML systems on GCP and managing GPU resources for enterprise-scale applications.

WORK EXPERIENCE

AI Engineer - Taishin Bank Jul 2025 - Present

- Developed speech recognition API using Breeze-ASR-25, replacing Whisper model to improve transcription accuracy for banking applications
- Built virtual human streaming system with load balancing and resource allocation, enabling multi-user concurrent access for customer service
- Implemented image generation API using Flux model with background removal and image overlay features
- Created video generation pipeline using Wan series models, extending output from 5 seconds to 30 seconds

AI Engineer - Shin Kong Financial Holding Aug 2024 - Jun 2025

- Deployed on-premise Whisper model replacing Azure STT for meeting transcription, achieving 84% reduction in batch processing time and 60-70% time savings in meeting documentation
- Developed customer service voice assistant using Llama-3.1-8B for Shin Kong Life Insurance, completing POC validation with management approval
- Built medical prescription recognition POC using Gemma-3-27B and OCR, achieving 70-80% accuracy with <30 seconds processing time per image
- Created insurance policy knowledge base with RAG pipeline using FAISS, indexing 60+ pages of policy documents
- Maintained Legal Compliance AI comparison model, resolving user-reported issues within 4-day average turnaround

Machine Learning Engineer - Sun-innovative Dec 2023 - Aug 2024

- Developed AI chatbot integrating GPT API with messaging platforms using RAG architecture and function calling for structured data queries
- Built voice meeting bot using Fast-Whisper for STT and PyAnnote for speaker diarization
- Implemented meeting summary generation, mind map visualization (Markmap), and one-click PPT generation (Python-pptx)
- Deployed services on GCP using Cloud Functions and Cloud Run with Firebase backend

Data Scientist (Intern) - Merkle Mar 2023 - Sep 2023

- Built automated data pipeline for L'Oréal integrating 10 data sources using GCP services (Dataflow, Cloud Functions, Airflow, BigQuery)
- Developed customer segmentation for KFC using K-means clustering and SVD++ for product preference prediction

SKILLS

LLM/SLM	Llama 3.1, Gemma-3, GPT API, LangChain, RAG, FAISS, Prompt Engineering
Speech & Vision	Breeze-ASR, Fast-Whisper, PyAnnote, OCR, Flux, Wan
Cloud	GCP, Azure
Programming	Python, SQL, Flask, FastAPI, Streamlit
Databases	Firebase, MySQL, Redis, BigQuery

EDUCATION

2021 - 2023 M.S. in Finance at **National Taipei University of Business**

Thesis: Predicting Financial Market Trends Using Sentiment Analysis and Trading Strategies

2016 - 2020 B.S. in Finance at **Ming Chuan University**

CERTIFICATIONS

Microsoft Azure Data Scientist Associate, Azure AI Engineer Associate

Google Google Analytics Certification