Yueh-Po Peng

Taipei, Taiwan | yuehpo.peng@gmail.com | linkedin.com/in/yueh-po-peng | github.com/y10ab1 | Google Scholar

Experience

Senior Machine Learning Engineer Jul 2025 – Present
Gamania Digital Entertainment Taipei, Taiwan
AI Manga & Webtoon Creation Platform (Muztrix)

- Led development of Muztrix, a public AI platform where a conversational agent collaborates with creators through multi-turn dialogue (Ask/Agent modes, tool calling, web search, file uploads) to develop storylines, then orchestrates generation of characters, locations, storyboards, and page or vertical-scroll manga.
- Designed a retrieval-augmented storyboard system with reference-image binding and versioned image editing to keep characters and locations visually consistent across panels.
- Integrated multiple generation models for images, line art, image-to-3D, and music, and built an AI game studio where coding agents create and publish web games inside isolated cloud sandboxes.
- Shipped a full-stack production service (Next.js, FastAPI, PocketBase on GCP Cloud Run) with CI/CD, monitoring, usage-based credits, OAuth, and 4-language i18n; reduced manga chapter prototyping from days to under 30 minutes.

AI Vtuber & Voice

- Trained voice cloning models enabling AI virtual characters to cover songs in target speaker voices with ~30 minutes of training data.
- Integrated Live2D avatar animation with LLM-driven persona dialogue and voice cloning, and designed expressive TTS voices for game characters.

Visiting Researcher (Contract) Jun 2025 – Oct 2025
Sony Computer Science Laboratories (Sony CSL) Tokyo, Japan (Hybrid)
Research Topics: Music Information Retrieval, Automatic Music Transcription

- Proposed VioPTT, a lightweight cascade model for violin technique-aware transcription with a novel synthetic dataset (MOSA-VPT). Published at ICASSP 2026 **[1]**.
- Investigated transfer learning for violin AMT, showing training from scratch on instrument-specific data can match fine-tuned piano-pretrained models. Published at ISMIR 2025 Late Breaking/Demo **[2]**.

AI Engineer Oct 2024 – May 2025
Gate.io Taipei, Taiwan (Remote)

- Developed a Text-to-SQL AI agent enabling non-technical teams to access internal data, boosting query efficiency by 20%.
- Developed a fund flows anomaly detection system with LLMs and tree-based models, enhancing financial security.

Machine Learning Engineer (Research) Mar 2022 – Oct 2024
Academia Sinica Taipei, Taiwan
Institute of Information Science · MCTLAB · Supervisor: Dr. Li Su
Research Topics: Self-Supervised Learning, Medical Imaging

- Designed a Transformer-based self-supervised learning pipeline for large-scale 4D fMRI, cutting GPU memory footprint by 77% through architectural redesign.
- Built and ran multi-node distributed training on TWCC HPC (Slurm) for high-resolution volumetric brain data.
- Developed a whole-brain feature selection method for fMRI decoding **[5]**, later extended into transferable whole-brain representations **[3]**.

AI Engineer Intern Mar 2023 – Jul 2024
Tomofun - World's leading pet technology company Taipei, Taiwan
Research Topics: Computer Vision, Large Language Models, Multimodal Learning

- Developed an automatic short music video generation system for daily pet clips.
- Fine-tuned visual language models (e.g., BLIP), achieving a 20.6% improvement in visual question answering.
- Enhanced LLaVA image inference speed by 250% with only a 3% accuracy reduction.
- Developed APIs for visual language models using llama.cpp/ollama for image-caption pair datasets.

Education

National Taiwan University Feb 2023 – Jun 2024
Taipei, Taiwan

- M.S. in Data Science | GPA: 4.09/4.30
- Thesis topic: Whole-Brain Feature Selection Methods for Decoding from fMRI Data

National Taiwan University Sep 2019 – Jan 2022
Taipei, Taiwan

- B.S. in Computer Science and Information Engineering (CSIE)

Research & Projects

Guitar Effect Removal Collaboration with Positive Grid ML Team

- Proposed a two-stage method to remove distortion effects from guitar recordings using Positive Grid VST plugins.
- Achieved 20% higher audio quality than the best baseline, rated by 26 professional guitarists.
- Published in DAFx 2024 **[4]**.

Whole Brain fMRI Feature Selection

- Proposed a two-stage method to extract fMRI features and predict musical pitch.
- Demonstrated 2-fold improvement over ROI-based feature selection in fMRI-music analysis.
- Published in ICASSP 2023 **[5]**.

Publications

[1] Wang, T. K.*, **Peng, Y. P.***, Su, L., & Cheung, V. K. M. "VioPTT: Violin Technique-Aware Transcription from Synthetic Data Augmentation," Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing 2026 (ICASSP'26). **(* equally contributed)** **Paper**
[2] **Peng, Y. P.***, Wang, T. K.*, Su, L., & Cheung, V. K. M. "Is Transfer Learning Necessary for Violin Transcription?," Int. Society for Music Information Retrieval Conf. 2025 (ISMIR'25) - Late Breaking/Demo. **(* equally contributed)** **Paper**
[3] **Peng, Y. P.**, Cheung, V. K. M., & Su, L. "Whole-Brain Transferable Representations from Large-Scale fMRI Data Improve Task-Evoked Brain Activity Decoding," arXiv preprint arXiv:2507.22378, 2025. **Paper**
[4] Lee, Y. S.*, **Peng, Y. P.***, Wu, J. T., Cheng, M., Su, L., & Yang, Y. H. "Distortion Recovery: A Two-Stage Method for Guitar Effect Removal," Proc. Int. Conf. Digital Audio Effects 2024 (DAFx'24). **(* equally contributed)** **Paper** | **Demo**
[5] Cheung, V. K.*, **Peng, Y. P.***, Lin, J. H., & Su, L. "Decoding Musical Pitch from Human Brain Activity with Automatic Voxel-Wise Whole-Brain FMRI Feature Selection," Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing 2023 (ICASSP'23). **(* equally contributed)** **Paper**

Skills

- **Languages/Frameworks**: Python, PyTorch, TensorFlow, Hugging Face, LangChain, LlamaIndex, Pandas, Scikit-learn, TypeScript, JavaScript, React, Next.js, FastAPI, Go, Slurm, C++, C, Linux.
- **Cloud/Infra**: GCP (Cloud Run, Vertex AI), AWS, Cloudflare Workers, TWCC HPC, Docker, PocketBase, Git, CI/CD.
- **Skillset**: LLM Agents, Tool Calling, RAG, Vision-Language Models, Image/Audio Generation, Prompt Engineering, Text-to-SQL, Voice Cloning, Self-Supervised Learning, Computer Vision, Music Information Retrieval, Multimodal Learning, Distributed Training.