

Ting-Yu Tsai

(no sponsorship required, permanent resident)



Summary

Ph.D. candidate in Computer Science with experience building multimodal LLM, machine learning, computer vision, and Python-based software systems for real-world applications. Developed end-to-end deep learning pipelines, evaluation workflows, and full-stack AI tools for medical imaging, video understanding, and intelligent decision-support systems. Brings strong hands-on experience in Python, PyTorch, TensorFlow, multimodal LLM inference, model evaluation, and software development, with a track record of translating research ideas into practical and scalable solutions.

Education

- **University at Albany - SUNY** *Albany, NY, USA*
Ph.D. in Computer Science *Aug. 2023 – Present*
- **National Taiwan University** *Taipei, Taiwan*
M.S. in Biomedical Electronics and Bioinformatics *Aug. 2019 – Jul. 2021*
- **Chung Shan Medical University** *Taichung, Taiwan*
B.S. in Medical Informatics *Aug. 2015 – Jul. 2019*

Skills

- **Languages:** Python, C/C++, JavaScript, PHP, MATLAB, SQL, HTML/CSS, VBA
- **AI/ML:** PyTorch, TensorFlow, scikit-learn, NumPy, Pandas, machine learning, deep learning, multimodal LLMs, multimodal AI, computer vision, model evaluation, inference benchmarking, medical image analysis
- **Backend / Data:** Node.js, MongoDB, MySQL, full-stack application development, workflow automation, data processing
- **Tools:** Git, Linux, Docker, OpenSSL

Experience

- **Research Assistant** — University at Albany - SUNY *Albany, NY*
Jan. 2025 – May 2025
 - Adapted the Quamba2 pipeline for Mamba-2 model quantization and reproduced benchmark results across model scales, enabling systematic evaluation of inference efficiency under quantization.
 - Built evaluation workflows to profile latency and memory usage and analyze activation behavior, state-space dynamics, and performance trade-offs across model variants.
- **Research Assistant** — University at Albany - SUNY *Albany, NY*
Jan. 2024 – Dec. 2024
 - Developed and benchmarked Python/PyTorch deep learning pipelines for cardiovascular image segmentation across challenging medical imaging datasets.
 - Improved model robustness and generalization by integrating Sharpness-Aware Minimization (SAM) and uncertainty-aware loss design into the training pipeline.
 - Designed and executed systematic evaluation workflows to optimize segmentation performance and support reliable medical image analysis.
- **Process Engineer** — Taiwan Semiconductor Manufacturing Co., Ltd. (TSMC) *Tainan, Taiwan*
Nov. 2021 – Jun. 2022
 - Supported semiconductor manufacturing operations by monitoring equipment performance, identifying process issues, and contributing to production reliability and efficiency.
 - Conducted risk assessments and collaborated with cross-functional teams to resolve equipment- and process-related issues while supporting quality and safety standards.
- **Software Engineer Intern** — Test Research Inc. *Taipei, Taiwan*
Aug. 2019 – Jul. 2021
 - Developed and optimized a deep learning model for X-ray image super-resolution using Python, TensorFlow, and TensorLayer.
 - Improved image quality and enlargement efficiency through model design, experimentation, and performance tuning.
 - Benchmarked the system against five classical AI models and four interpolation methods, demonstrating competitive quality and runtime performance.
- **Software Engineer Intern** — Chung Shan Medical University Hospital *Taichung, Taiwan*
Jul. 2018 – Aug. 2018
 - Developed a Python-based automated organ contouring system for head-and-neck CT images, building an end-to-end workflow for image conversion, organ segmentation, and RS file generation.

- Applied GAN-based deep learning to improve workflow consistency in radiation oncology image processing.
- Reduced organ selection and contouring time by 70%, improving efficiency for clinical staff.

• **Developer** — Chung Shan Medical University / Chung Shan Medical University Hospital

Taichung, Taiwan

Apr. 2016 – Sep. 2018

- Designed and developed database-driven web platforms using PHP, JavaScript, HTML, CSS, MySQL, and OpenSSL to support document management, review workflows, admissions outreach, and administrative processes.
- Built role-specific workflow and automation features that improved process visibility, reduced manual workload, and supported institution-specific operational requirements.
- Automated internship placement and academic document workflows, reducing processing time from approximately one week to under one minute for selected tasks.

Selected Projects

• **ReDiPrune: Efficient Multimodal LLM Inference** [1]

Developed ReDiPrune, a training-free token pruning framework for multimodal LLMs that improves inference efficiency by pruning visual tokens before the vision-language projector. Designed a lightweight relevance-diversity scoring method to preserve query-relevant and non-redundant information, enabling plug-and-play integration into existing multimodal LLM pipelines without retraining or architectural changes. (2026)

• **IntelliCardiac: End-to-End AI Platform for Clinical Decision Support** [2]

Built an end-to-end full-stack AI platform for cardiac MRI segmentation, five-class disease classification, and real-time result visualization. Developed model pipelines, evaluation workflows, and deployment-oriented application components using Python, PyTorch, JavaScript, Node.js, and MongoDB; the segmentation module achieved 92.6% overall accuracy and the two-stage classification pipeline reached 98% accuracy. (2025)

• **Rat Seizure Video Analysis Framework** [3]

Developed a deep learning and computer vision framework for automated rat seizure detection and stage classification from video. Built the video analysis pipeline to support scalable and reproducible behavioral monitoring for epilepsy research. (2026)

• **UU-Mamba: Robust Cardiovascular Image Segmentation** [4]

Developed and benchmarked a cardiovascular image segmentation framework that integrates Sharpness-Aware Minimization and an uncertainty-aware composite loss to improve robustness and generalization. Achieved state-of-the-art performance on the ACDC, ImageCAS, and Aorta datasets. (2024)

• **Automatic Image Exposure Classification and Enhancement System**

Built a deep learning-based system for automatic image exposure classification and enhancement, achieving 0.91 classification accuracy and improving image quality to a level comparable to camera-captured results. (2021)

Selected Publications

- [1] A. Yu, **Ting Yu Tsai**, Z. Zhang, W. Lu, F. X. .-. Ye, and M.-C. Chang, “ReDiPrune: Relevance-Diversity Pre-Projection Token Pruning for Efficient Multimodal LLMs,” *arXiv preprint arXiv:2603.24680*, 2026.
- [2] **Ting Yu Tsai**, A. Yu, et al., “IntelliCardiac: An Intelligent Platform for Cardiac Image Segmentation and Classification,” in *2025 IEEE 8th International Conference on Multimedia Information Processing and Retrieval (MIPR)*, 2025, pp. 109–115.
- [3] **Ting Yu Tsai**, A. Yu, et al., “RatSeizure: A Benchmark and Saliency-Context Transformer for Rat Seizure Localization,” *arXiv preprint arXiv:2603.26780*, 2026.
- [4] **Ting Yu Tsai**, L. Lin, S. Hu, M.-C. Chang, H. Zhu, and X. Wang, “UU-Mamba: Uncertainty-aware U-Mamba for Cardiac Image Segmentation,” in *2024 IEEE 7th International Conference on Multimedia Information Processing and Retrieval (MIPR)*, 2024, pp. 267–273.
- [5] X. Wang, **Ting Yu Tsai**, et al., “Spotting the Fakes: A Deep Dive into GAN-Generated Face Detection,” *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 21, no. 7, pp. 1–24, Jul. 2025.

Selected Honors

First Prize and Best Popularity Award, Taipei Medical University Venture Capital Investment Contest, 2017

First Prize, International Conference on Medical and Health Informatics, 2017

Gold Prize, Korea International Women’s Invention Exposition (KIWIE), 2013