Te-Yi "Dexter" Kan

⚲ Los Angeles, CA   ▯ (818) 309-0203   ✉ teyikan@usc.edu

in dexterkan   🌐 Personal Website   🎓 Google Scholar

Research Areas

Privacy-Preserving ML, Neural Architecture Search (NAS), Lightweight & Efficient ML, Edge Computing & Task Scheduling, Computer Vision, Large Language Models (LLMs), Wireless Networking & Communications.

Technical Skills

Programming Languages:	Python, C/C++, Java, Bash, Matlab, R
Frameworks & Libraries:	Pandas, PyTorch, TensorFlow, Transformers, OpenStack, Docker, Kubernetes

Education

University of Southern California - *Ph.D. in ECE*   Aug. 2022 - May 2027 (expected)
- Dissertation Topic: Towards Efficient and Privacy-Preserving Machine Learning for Resource-Constrained Devices
- Coursework: Stochastic Network Optimization, Operating Systems, Computer Vision, Artificial Intelligence, Reinforcement Learning & Large Language Models

University of California, Los Angeles - *M.S. in ECE*   Sep. 2019 - Dec. 2020
- Coursework: Convex Optimization, Neural Networks and Deep Learning, Reinforcement Learning

National Taiwan University - *B.S. in EE*   Sep. 2014 - Jan. 2019
- Coursework: Wireless and Mobile Networking, Wireless Communications, Network Planning and Management

Selective Publications

[1] **Te-Yi Kan**, Konstantinos Psounis, "Low-Latency Private ML inference for Vision Tasks in Distributed Environments", *IEEE Transactions on Dependable and Secure Computing*, 2025, submitted. (TDSC-25)

[2] **Te-Yi Kan**, Konstantinos Psounis, "Online distributed offloading of time-sensitive vehicular tasks in edge-cloud systems", *IEEE Transactions on Vehicular Technology*, 2025. (TVT-25) | ✇

[3] **Te-Yi Kan**, Ronald Y. Chang, Feng-Tsun Chien, "Hybrid Intelligent Reflecting Surface and Classical Relay Assisted Multiuser MISO Systems", *IEEE Transactions on Vehicular Technology*, 2023. (TVT-23) | ✇

[4] Tz-Wei Mo, Ronald Y. Chang, **Te-Yi Kan**, "DeepMCTS: Deep Reinforcement Learning Assisted Monte Carlo Tree Search for MIMO Detection", *2022 IEEE 95th Vehicular Technology Conference: (VTC2022-Spring)*, 2022. (VTC-22) | ✇

[5] **Te-Yi Kan**, Ronald Y. Chang, Feng-Tsun Chien, "Intelligent Reflecting Surfaces and Classical Relays: Coexistence and Co-Design", *2021 IEEE GlobeCom Workshops (GC Wkshps)*, 2021.(GC-21). | ✇

[6] **Te-Yi Kan**, Yao Chiang, Hung-Yu Wei, "QoS-aware Mobile Edge Computing System: Multi-server Multi-user Scenario", *2018 IEEE GlobeCom Workshops (GC Wkshps)*, 2018. (GC-18) ✇

[7] **Te-Yi Kan**, Yao Chiang, Hung-Yu Wei, "QoS-aware Fog Computing System: Load Distribution and Task Offloading", *The 15th IEEE Vehicular Technology Society Asia Pacific Wireless Communications Symposium (VTS APWCS)*, 2018. (APWCS-2018) | ✇

[8] **Te-Yi Kan**, Yao Chiang, Hung-Yu Wei, "Task Offloading and Resource Allocation in Mobile-Edge Computing System", *2018 27th Wireless and Optical Communication Conference (WOCC)*, 2018. (WOCC-18) | ✇

Research Projects

Efficient Hardware-aware Neural Architecture Search for Mobile Devices [*ICML-26*]   Jun. 2025 - Present
- Designed a two-stage NAS framework combining cloud-based search with edge-based fine-tuning, enabling high-performance personalized models under strict latency and memory constraints.
- Conducted extensive experiments on CIFAR-100, ImageNet, and Freiburg datasets, with cross-dataset comparisons to evaluate effectiveness under varying edge-cloud distribution shifts.

Privacy-Preserving ML Systems [TDSC-25]   Jun. 2024 - Present
- Built a privacy-preserving inference pipeline with sensitive object detection, task scheduling, and obfuscation modules for vision tasks.

- Leveraged distributed remote servers to balance utility and privacy.
- Achieved $\sim$30% higher task utility at 25 fps without compromising user privacy.

Online Distributed Offloading Mechanism in Vehicular Networks [*TVT-24*] Feb. 2023 - May 2024
- Developed TODORA, a threshold-based online distributed offloading and resource allocation mechanism for real-time vehicular ML tasks.
- Reduced end-to-end task duration (execution delay + communication overhead) by optimizing offloading and allocation under dynamic edge–cloud environments.
- Demonstrated superior performance over state-of-the-art baselines via extensive simulations.

Intelligent Reflecting Surface (IRS) [*TVT-23, GC-21*] Jan. 2021 - May 2022
- Proposed a multiuser downlink MISO system assisted by a full-duplex IRS and half-duplex DF relay.
- Designed an alternating-optimization algorithm to jointly optimize active BS/relay beamforming and IRS passive beamforming for maximum sum-rate.
- Demonstrated significant throughput gains and analyzed tradeoffs in joint beamforming design.

Multiple-input multiple-output (MIMO) Detection [*VTC-22*] Oct. 2020 - Apr. 2021
- Devised a deep reinforcement learning enhanced Monte Carlo Tree Search (MCTS) framework for MIMO detection.
- Demonstrated notable performance and complexity improvements over conventional MCTS across diverse channel conditions.

Multi-access Edge Computing (MEC) Systems [*GC-18, APWCS-18, WOCC-18*] Sep. 2017 - Aug. 2018
- Developed QoS-aware task offloading strategies for MEC systems in single- and multi-server settings.
- Proposed a two-stage algorithm optimizing task offloading, resource allocation, and load distribution.
- Validated superior latency reduction and QoS performance against multiple benchmarks.

Work Experience

Member, Academia Sinica-MediaTek Collaborative Project Oct. 2021 - May 2022
- Designed an AI-enabled hybrid RIS/relay system to improve energy efficiency and system sum-rate.
- Analyzed system performance and tradeoffs in mmWave and THz frequency bands.

Summer Intern, Foxconn Technology Group Jun. 2017 - Aug. 2017
- Implemented a commercial MEC infrastructure to enable lifecycle management of MEC services.
- Contributed to the development of an MEC-powered face recognition system for access control and attendance management, using OpenStack for deployment.

Selective Course Projects

Schematic-to-Netlist Translation using RL-Enhanced VLM Pipeline Fall 2025
- Developed an RL-enhanced VLM pipeline to translate circuit schematic diagrams into SPICE-compatible netlists.
- Trained Qwen with LoRA for schematic understanding, then used GRPO with syntax and functional correctness as reward signals to iteratively refine translation quality.

Generative Adversarial Imitation Learning (GAIL) Spring 2020
- Implemented a scalable imitation learning algorithm in Python for high-dimensional environments.
- Achieved superior performance over Behavior Cloning on OpenAI Gym tasks (Pendulum, Cartpole).

Image Generation Methods for Cataract Surgery Winter 2020
- Applied U-Net and U-Net+WGAN frameworks to segment cataract surgery images of eye structures.
- Benchmarked models using pixel accuracy, mean IoU, and F1 score, with comparisons against self-labeled data.

IEEE 802.15.4 Design Challenge Fall 2017
- Designed duty cycle–adaptive Machine-to-Machine (M2M) communication protocols for wireless networks.
- Implemented protocols on FCM2401 modules using AutoNet SDK.

Functionally Reduced And-Inverter Graph (FRAIG) Fall 2016
- Built a C-based FRAIG circuit representation with data structures for detecting functionally equivalent pairs.
- Integrated hashing, Boolean simulations, and SAT solvers to identify equivalences in digital circuits.