

Feng-Ting Liao

MediaTek Research | Taipei, Taiwan
[ftliao.github.io](https://github.com/ftliao) | fengting.liao19@gmail.com | +886-972331269

Research Scientist and Engineer specializing in LLMs, Deep RL for chip design, and AI. Co-creator of Breeze and BreeXe, with proven expertise in productization, interdisciplinary collaboration, and research publications.

WORK EXPERIENCE

Senior Research Scientist | MediaTek Research, MediaTek Nov 2020 - Present, Taipei, Taiwan

- **Project lead of deep reinforcement learning for floor planning in chip design.**
 - Optimized floor-planning process in chip top design (6 weeks down to 6 hours). Formulated, prototyped, and delivered deep reinforcement learning agents based on graph neural network and Transformer.
 - Coordinated cross-departmental collaboration to integrate and evaluate AI agents on chip design cases; cross-institutional collaboration for building shaping and routing tools for RL environments.
- **Co-creator of Breeze-7B/BreeXe-8x7B**, SOTA Traditional Chinese LLMs with **>6k downloads/month**.
 - **Productized BreeXe-8x7B via cloud API and on-premises deployment** to the Traditional Chinese market through aligning diverse stakeholders including internal leadership, legal, marketing, and product and external system integrators and IT departments of corporations within the ecosystem.
 - Core-contributor to key stages of LLM development for Breeze-7B and BreeXe-8x7B, including data collection, pre-training and post-training processes, evaluation, marketing, and productization.
- **Research publications at ICML, ACL, NAACL, and ASRU** in Large Language Model, Diffusion Models, Meta-Learning, Natural-Language-Processing, and Speech Recognition.
- **Advanced LLM evaluations** through open-sourcing TCEval, the first comprehensive language model benchmark in Traditional Chinese; RAD-Bench, the first benchmark on retrieval augmented dialogues.
- **Advanced cross-modal application of LLMs** via proposing Generative Fusion Decoding, an algorithmic framework for shallow fusing speech recognition model and LLM without re-training.
- Mentor for interns on algorithm prototyping and research publications; coordinator of staff hiring, cross-departmental reading groups, and lab compute infrastructure.

Research Engineer in Computer Vision | Umbo CV Mar 2019 - Feb 2020, Taipei, Taiwan

- Led a 90% cost reduction and 350% service speed boost for cloud cameras through algorithmic optimization; architected a vehicle detection API for real-time surveillance with cross-functional teams.

Data Science Intern | Burberry Dec 2018, London, UK

- Enhanced the digital retail experience for 12M customers during Winter 2018 sales by designing and deploying a trending algorithm using the Mann-Whitney U test.

Postdoctoral Research Assistant | University of Oxford Jan - Dec, 2018, Oxford, UK

DPhil Researcher | University of Oxford Oct 2013 - Jan 2018, Oxford, UK

- Core-contributor to the first result of the world's largest dark matter detector in direct detection.
- Led the Oxford team in delivering £50K worth of sensors and investigated applying machine learning to dark matter research. Designed a state-of-the-art monitoring system and collaborated with ~20 researchers on the detector's time projection chamber.

BACKGROUND & SKILLS

ML areas: Large Language Model, Deep Reinforcement Learning, Speech AI, Meta-Learning, Diffusion Models, Generative Models, NLP, Computer Vision

Programming skills: Python, C++, Shell (proficient); SQL (basic)

Technologies: PyTorch, JAX, CUDA, NumPy, Pandas, Docker, Kubernetes, Git, Jenkins, AWS, GCP, Tableau

EDUCATION

DPhil in Particle Physics, University of Oxford 2013 - 2018

B.Sc. in Electrophysics, National Chiao Tung University 2008 - 2012