TING-YU DAI

www.funnyengineer.com
2100 E. Yager Ln. Apt 2405, Austin, TX, 78754
(+1)5128319769 ⋄ funnyengineer@utexas.edu ⋄ FunnyEngineer ⋄ Ting-Yu Dai

EDUCATION

Ph.D. Candidate in Sustainable System, *University of Texas at Austin* — Austin, TX *2021 - Present*
Advisors: Prof. Dev Niyogi & Prof. Zoltan Nagy

MSc in Computer-Aided Engineering, *National Taiwan University* — Taipei, Taiwan *2019 - 2021*

BS in Civil Engineering, *National Chiao Tung University* — Hsinchu, Taiwan *2015 - 2019*

RESEARCH INTERESTS

Climate Change, Machine Learning, Building Energy Modeling, Geospatial Data, Self-supervised Learning

SKILLS

Programming	Python, TypeScript, JavaScript, Dart (Flutter), C++, Java, MATLAB, C#
Machine learning framework	Pytorch, Tensorflow, Detectron2, Scikit-learn, AWS Lex, Darts
Software & Tools	AWS, Firebase, PostgreSQL, Unity, MySQL, MongoDB, Linux, Git, Docker, Tableau

RESEARCH EXPERIENCE

PrecipDiff: Leveraging diffusion models to enhance satellite-based precipitation – *Diffusion, Downscaling*

[**AAAI 2025**] The study introduces the first diffusion model for correcting discrepancies among precipitation data, enabling downscaling of satellite estimates from 10 km to 1 km resolution. Experiments in Seattle indicate notable improvements in accuracy and spatial detail, highlighting the efficacy of a computer vision-based approach to enhance precipitation data from satellites. *Feb. 2025*

CityTFT: Temporal Fusion Transformer for Urban Building Energy Modeling – *Transformer, Energy*
[**NeurIPS 2023**] Established a temporal fusion transformer to model urban energy demands as a surrogate model for traditional physic-based UBEM methods. CityTFT reached **40 times** faster to simulate compared to the physics-based model and **6 times** more accurately compared to classic RNN and transformers while predicting in an unseen climate dynamic. (F1 score of **99.98 %** while RMSE of loads of **13.57 kWh**.) *Oct. 2023*

Analyzing the impact of COVID-19 on the electricity demand in Austin, TX using an ensemble-model based counterfactual and 400,000 smart meters – *Ensemble Model, Social Science, Building Energy*
[**Urban Computational Science**] Applied a large-scale private smart meter electricity demand data from **the City of Austin**, combined with publicly available environmental data, and develops an ensemble regression model for long-term daily electricity demand prediction. *Dec. 2022*

Generating High-Resolution PM2.5 using a Two-stage Machine Learning Approach with Low-Cost Air Quality Sensors and Satellite Observations – *Data Fusion, Air Quality, Remote Sensing* **[REF]**
[**AGU2022 Oral**] Developed a two-stage machine learning method to create **a ground-level PM2.5 grid dataset** by calibrating LCS and using the calibrated PM2.5 to fuse with HRRR(Meteorological data) and AOD values. *Dec. 2022*

Modelling high-resolution rainfall extremes in a changing climate – *Self-Attention, Rainfall Extremes* **[REF]**
[**MSc Thesis**][**EGU2021**] Implemented an ML-based approach to bridge climate reanalysis data and local rainfall statistics and predicted future rainfall patterns based on future climate. *Apr. 2021*

EXPERIENCE

Fujitsu Research of America San Jose, California
Research Intern – Diffusion, Donwscaling *May. 2024 - Dec. 2024*

- Working in Converging Technology Lab for a digital climate project
- Develope a downscaling method using diffusion model and apply on operational precipitation dataset

NASA, Universities Space Research Association (USRA) Huntsville, Alabama
Research Intern – machine learning, air quality, geospatial data *May. 2022 - Aug. 2022*

- Working with the **NASA Marshall Space Flight Center** research team for a Citizen Science Project.

- Utilized PurpleAir sensor in San Francisco and Los Angeles and developed a machine learning model to calibrate the LCS measurements with the federal equivalent methods which **decrease the MSE from 6.38 to 0.11**.
- Designed a data fusion method to merge meteorology and AOD data into the ground-level PM2.5 concentration and generated an urban gridded PM2.5 dataset in both SF and LA area that contains **over 134 million data points.**