Ye Tao
(585)755-8637 • yt371@rutgers.edu
https://www.linkedin.com/in/ye-tao-12430a167/

Qualifications

- Experience in building analytical frameworks including data processing, data mining, and quantitative analysis
- Background in machine learning and statistical analysis under the framework of differential privacy
- Fluent in Matlab and Python including libraries such as sqlalchemy, tensorflow, pytorch, langchain, openai
- Leadership experience as lab manager for data management and communication between collaborators

Education

Rutgers University
PhD candidate in Electrical and Computer Engineering *Sept 2020 – Oct 2025 (expected)*
- Research interests: Differential privacy, machine learning, deep learning, statistical learning, optimization

University of Rochester
Master of Science in Electrical and Computer Engineering *Sept 2017 – May 2019*

Research & Project Experience

Rutgers University
Federated Learning with Differential Privacy Using BrainLM for Preprocessing *Nov 2024 – present*
- Use the brainLM foundation model to enhance federated differential privacy model training with 4D fMRI data
- Fine-tune brainLM, built on a transformer-based masked autoencoder architecture, using more public datasets
- Extract useful features from trained transformer encoder, utilizing its ability to capture complex data patterns
- Utilize the low-dimensional features for training downstream differential privacy models in a federated setting

Federated Privacy-Preserving Visualization *Sept 2024 – Nov 2024*
- Integrate federated learning with differential privacy for privacy-preserving data visualizations
- Propose a framework with visual representations, privacy mechanisms, integration approaches, and evaluations
- Identify potential challenges in combining federated learning and differential privacy and propose solutions
- Demonstrate the feasibility of our approach with examples such as scatterplots and correlation visualizations

Differentially Private (DP) Distribution Estimation Using Functional Approximation *Nov 2023 – May 2024*
- Develop a novel DP distribution estimation method inspired by functional analysis and functional mechanism
- Project empirical distribution into a predefined polynomial space and approximate using Legendre polynomials
- Achieve a DP empirical CDF by protecting coefficients of Legendre polynomials through Gaussian mechanism
- Outperform existing privacy-preserving methods in federated settings and scenarios involving streaming data

Privacy-preserving visualization of brain functional network connectivity *Sept 2022 – Nov 2023*
- Introduce novel methods for privacy-preserving visualization on neuroimaging data like functional connectivity
- Propose an end-to-end differentially private workflow for connectogram visualization on group comparison
- Enhance visualization quality through various strategies such as pre- and post-processing approaches
- Show that differentially private mechanisms have different optimal composition strategies

University of Rochester
Data mining for brain analysis in Mild Cognitive Impairment (MCI) *Feb 2019 – May 2020*
- Design pipeline specifically for resting-state fMRI data preprocessing using FSL and Freesurfer software
- Implement Independent Component Analysis (ICA) to identify 14 brain networks from fMRI data
- Select 9 out of 14 networks that exhibited great correspondence with results from other literatures
- Apply general linear model and dual regression to extract networks with regions showing group difference
- Identify activated brain regions related with visual tasks to serve as potential biomarkers for MCI analysis

Work Experience

LinkedIn Corporation

Applied Research Intern *May 2024 – Aug 2024*

- Construct an end-to-end model to integrate long-term and short-term rewards for ad recommendations
- Define long-term reward with return-to-go from RL and use SQL to process billions of records to generate it
- Utilize long-term rewards as multi-head labels or as transformer input for effective model training
- Boost user pCTR by leveraging long-term rewards for additional information over immediate rewards

Applied Research Intern *May 2023 – Aug 2023*

- Design an end-to-end GAI system to retrieve and process structured as well as unstructured text data
- Identify the type and stage of the task and decide the next step through different agents in the system
- Perform tasks like writing SQL scripts, retrieving info from various databases, generating reports, etc.
- Enhance productivity in sales and marketing through various functionalities implemented in the system

Applied Research Data Science Intern *May 2022 – Aug 2022*

- Build a multi-task (representation) learning framework for various prediction tasks in Go-To-Market domain
- Learn universal patterns common to multiple datasets and distinctive patterns specific to each dataset
- Train customized representations with MTRL utilizing pre-trained representations from other domains
- Fill the gap between different stages in the marketing and sales process via the overall framework

Rutgers University

Teaching Assistant of Machine Learning course *Aug 2023 – Dec 2023*

- Design weekly recitation sessions for a comprehensive review of lecture content and create exercises
- Prepare and host in-class programming exercises, guide term projects, and answer questions on Piazza

Teaching Assistant of Linear Systems and Signals course *Aug 2020 – Dec 2020 & Aug 2023 – Dec 2023*

- Provide feedback on student learning progress with course instructor during weekly meetings
- Prepare weekly recitation lectures for classroom knowledge review and exercises of typical questions
- Design course assignments and quizzes with corresponding solutions and hold weekly office hours

University of Rochester

Information Analyst and Data Manager at CogT lab *Aug 2019 – June 2020*

- Communicate with project coordinators to ensure efficient information exchange for data management
- Monitor incoming data from multiple sources such as paper questionnaires, online tests, and MRI scanning
- Construct and manage database on FileMaker and REDCap for lab data storage and retrieval
- Coordinate neuroimaging data transfer within department and among international research institutions

Research Assistant at CogT lab *Feb 2019 – May 2019*

- Compose Matlab and Python scripts for neuroimaging data collection, organization, and preprocessing
- Apply signal processing techniques on neuroimaging modalities, e.g., fMRI, structural MRI, DTI
- Analyze neuroimaging data using statistics and graph theory methods on structural and functional MRI data

Publication

- **Ye Tao**, Anand D. Sarwate, 2025. Differentially Private Distribution Estimation Using Functional Approximation. (Accepted for ICASSP 2025)
- **Ye Tao**, Anand D. Sarwate, Sandeep Panta, Sergei Plis, and Vince D. Calhoun, 2024. Federated Privacy-Preserving Visualization: A Vision Paper. (Accepted for IEEE Big Data 2024)
- **Tao, Y.**, Sarwate, A. D., Panta, S., Plis, S., & Calhoun, V. (2024). Privacy-Preserving Visualization of Brain Functional Connectivity. *bioRxiv*, 2024-10.
- **Tao, Y.**, Sarwate, A. D., Panta, S., Plis, S., & Calhoun, V. D. (2024, May). Privacy-Preserving Visualization of Brain Functional Network Connectivity. In *2024 IEEE International Symposium on Biomedical Imaging (ISBI)* (pp. 1-5). IEEE.
- **Y. Tao**, A. Chihoub, A. D. Sarwate, S. Panta, and V. Calhoun, "Privacy-Preserving Visualization of Functional Network Connectivity," in International Conference of the IEEE Engineering in Medicine and Biology Society, Glasgow, Scotland, UK, 2022.

- Lin, F.V., **Tao, Y.**, Chen, Q., Anthony, M., et al., 2020. Processing speed and attention training modifies autonomic flexibility: A mechanistic intervention study. NeuroImage, 213, p.116730.

Skill

- Computer software: Microsoft Office suite, FileMaker, REDCap, Latex, SPSS, Gephi
- Programming languages: Proficient in Matlab, Python; familiar with SQL, HTML
- Data management skills in monitoring data processing pipeline, coordinating project progress within team