Yenho Chen

(210) 639-3070 | yenho@gatech.edu | yenhochen.github.io | U.S.A. Citizen

Education

PhD Machine Learning, Georgia Tech, Atlanta, GA — *Aug. 2021 - May 2027 (Expected)*

Advisor: Prof. Christopher J. Rozell — GPA: 4.0/4.0

Thesis: Dynamics-Based Self-Supervised Learning for Physiological Time Series

Recipient of University Presidential Fellowship

MS Computer Science, Georgia Tech, Atlanta, GA — *Aug. 2021 - Dec. 2023*

Specialization in Visual Analytics — GPA: 4.0/4.0

BS Physics, University of Texas at Dallas, Richardson, TX — *Aug. 2015 - May 2019*

Magna Cum Laude, Recipient of University Academic Excellence Award with Distinction — GPA: 3.9/4.0

Selected Papers

Y. Chen*, M. A. Xu, J. Regh, C. J. Rozell, "Self-Supervised Dynamical System Representations For Physiological Time-Series," *Accepted to the Proceedings of the 43rd International Conference on Machine Learning (ICML), 2026.*

B. Ancelin*, **Y. Chen***, A. Falcon, P. Guan, C. Kaushik, N. Singh, B. Urcelay, "MANGO: Learning Disentangled Image Transformation Manifolds with Grouped Operators," *Sampling Theory and Applications (SampTA)*, 2025. (Oral)

Y. Chen*, N. Mudrik, A. S. Charles, C. J. Rozell, "Probabilistic Decomposed Linear Dynamical Systems for Robust Discovery of Latent Neural Dynamics," *In Advances in Neural Information Processing Systems (NeurIPS)*, 2024.

N. Mudrik*, **Y. Chen***, E. Yezerets, C. J. Rozell, A. S. Charles, "Decomposed Linear Dynamical Systems (dLDS) for learning the latent components of neural dynamics," *Journal of Machine Learning Research (JMLR)*, 2024.

***** denotes first author or first co-authors, equal contribution

Work Experience

Research Assistant — *2021 - Present*

Georgia Tech — *Atlanta, GA*

- Led research on self-supervised learning methods for health time-series, using mathematical tools from dynamical systems theory to improve label efficiency and transferability of the learned representations across low-data and large-scale single- and multi-sensor settings.
- Initiated collaborations with Google, Rice University, and Johns Hopkins, producing first-author publications at **ICML, NeurIPS, and JMLR**.

Accelerated Discovery Research Intern — *Summer 2025*

IBM Research — *San Jose, CA*

- Designed a contrastive learning approach that uses evolutionary information from short linear motifs to predict phenotypes of mutated CAR-T sequences, enabling in silico screening for high-cytotoxicity variants and reducing de novo protein design costs.
- Submitted **a patent** and **a manuscript** by developing the core method in collaboration with experimentalists from Stanford and Altos Labs.

Artificial Intelligence Research Intern — *Summer 2024*

Blackstone Real Estate — *New York City, NY*

- Reduced Blackstone's hotel demand forecasting error by 26% with a novel differentiable dropout that contextually removes noisy modalities.
- Engineered a data ingestion pipeline with 35% faster throughput by optimizing SQL queries to accelerate ML experimentation.

Applied Large Language Model Intern — *Spring 2024*

Travelers Insurance Inc. — *Atlanta, GA*

- Achieved a 3x speedup in LLM pretraining efficiency by designing a domain reweighting method based on inter-domain gradient alignment.
- Increased evaluation throughput 5x by standardizing benchmarking across many HuggingFace LLM and reweighting method combinations.

Skills

Research	Self-supervised & contrastive learning, dynamical systems theory & state-space models, Bayesian inference, probabilistic graphical models, sparse coding, multimodal modeling, representation learning, transfer learning, semi-supervised learning
Applications	Health time-series (ECG, PPG, EEG, HAR, wearables), large language model (LLM) pretraining, protein language models
Software	PyTorch, Tensorflow, Scikit-Learn, Numpy, Python, C++, SQL, Linux, Git, Slurm, AWS, Databricks, Snowflake

Selected Research

Self-Supervised Dynamical Systems Pretraining for Multi-Sensor Health Time-Series *2026*

Manuscript in Preparation *Georgia Tech*

- Developed a self-supervised pretraining framework for multi-sensor health time-series via a dynamics-based hybrid cross-reconstruction contrastive objective, with preliminary results showing 11% average error reduction in linear probe across 5 benchmark datasets.
- Derived theory on recovering transferable dynamical systems information from multi-sensor time-series datasets via a cross-reconstruction objective, motivating a novel observation process augmentation that isolates physiological dynamics from heterogeneous sensor configurations.

Self-Supervised Dynamical Systems Pretraining for Single-Sensor Health Time-Series *2025-2026*

Accepted at **ICML 2026** *Georgia Tech*

- Designed PULSE, a self-supervised framework for time-series that achieves a 12% error reduction across EEG, ECG, HAR, and PPG benchmarks in linear probe and semi-supervised settings, via a novel cross-reconstruction objective that isolates physiological dynamics from noise.
- Established the first theory relating cross-reconstruction with dynamical systems models of time-series to predict the efficacy of data augmentation strategies at recovering transferable information about the underlying physiological dynamics.

Sequence-Level CAR-T Phenotype Prediction via Soft-Contrastive Learning *2025-2026*

Submitted **a Patent Application** and manuscript to **CIBB 2026** *IBM Research*

- Extended CAR-T protein cytotoxicity prediction from motif- to sequence-level, achieving $R^2 = 0.705$ on held-out sequences, by designing a soft-contrastive objective that leverages BLASTp sequence similarities and consensus-region annotations.
- Outperformed ESM-2 by 32% in consensus-mutation classification, demonstrating higher sensitivity to biologically relevant point mutations.

Learning Disentangled Image Transformations with Grouped Manifold Operators *2024-2025*

Published at **SampTA 2025 (Oral)** *Georgia Tech*

- Developed MANGO, a manifold model that learns interpretable operators for controlling image transformations via a block-diagonal group structure, achieving a 60% improvement in out-of-distribution reconstruction over standard autoencoders.
- Reduced training time by 100x over the standard Manifold Autoencoder by replacing a 3-phase training procedure with a 1-phase approach.

Stable and Interpretable Neural Dynamics via Probabilistic dLDS *2023-2024*

Published at **NeurIPS 2024** *Georgia Tech*

- Developed a variational EM framework that estimates dLDS sparse codes robustly under dynamical noise in non-stationary systems, reducing the latent dynamics tracking error by 2x over the strongest baseline in simulation.
- Demonstrated discovery of emotion-aligned dynamic modes from LFP recordings that generalize to held-out data, where all prior methods fail.

Decomposing Linear Dynamical Systems (dLDS) for Interpreting Neural Dynamics *2022-2024*

Published at **JMLR 2024** *Georgia Tech*

- Introduced dLDS, an interpretable sparse dictionary learning framework over linear dynamical components supporting combinatorially more dynamic regimes than prior switching approaches for a fixed parameter budget.
- Demonstrated that dLDS disentangles independent subnetworks and identifies shared neural dynamics in C. elegans whole-brain recordings, solving a previously intractable task through the novel sparse operator formulation.

Selected Leadership

Bioengineering and Bioscience Unified Graduate Students (BBUGS) *2022 - Present*

Social Chair *Georgia Tech*

- Quadrupled event frequency (semesterly →monthly) for a 2K+ member community by directing a team of 5 and managing a $6K annual budget.
- Established two annual traditions by reviving the overnight rafting trip post-COVID and co-founding the Garden Gala (100+ attendees each).

Reviewer Service

Journal of Neural Engineering (2023), NeurIPS (2024,2025), ICML (2025, 2026), ICLR (2025, 2026)

Selected Awards

ICML Gold Reviewer Award *2026*

NeurIPS Top Reviewer Award *2025*

IBB "Above and Beyond" Trainee Award *2025*

GT Machine Learning Student Conference (2nd Place Poster) *2024*

Georgia Tech President's Fellowship *2021 - 2024*

NIH Academy Scholar Award *2020*