

**Xujin “Chris” Liu**  
New York City, NY (10016)

Email : chris.liu@nyu.com  
Mobile : (412) 315-9733

## SUMMARY

---

- PhD candidate in Electrical & Computer Engineering (computational neuroscience) with a research-engineering focus: building scalable datasets, preprocessing pipelines, and self-supervised pretraining workflows for neurophysiology (EEG/iEEG/MEG) and clinical prediction. Publications in *Nature* and *Nature Communications*; experience with large-project debugging and training infrastructure.

## EDUCATION

---

- **New York University** New York, NY  
*Ph.D. Candidate, Electrical and Computer Engineering; GPA: 3.8/4.0* Sept 2021 – Present
- **Carnegie Mellon University** Pittsburgh, PA  
*B.S., Electrical and Computer Engineering; Additional Major: Biomedical Engineering; GPA: 3.7/4.0 July 2017 – May 2021*

## SELECTED PROJECTS (RESEARCH ENGINEERING & ML)

---

- **Scaling EEG language models for seizure detection** New York, NY  
*Empirical scaling law* 2024
  - **Scaling law of EEG-language models:** First systematic study of how tokenization granularity (duration per token) interact with language model size (0.6B-4B) and find that scaling tokenization granularity is more reliable for improving cross-site generalization compared to scaling model size.
  - **Representation analysis:** Despite flattening 2D signal into 1D, permutation invariance is still displayed in model representation. Also, fine-grained tokenization produces qualitatively different representations compared to coarser ones.
  - **Publication:** Manuscript under review at ICML 2026
- **BEEG Pretraining dataset & BEEG Bench** New York, NY  
*Dataset / Pipeline Engineering* 2024 – Present
  - **BEEG: Unified neurophysiology corpus:** Collected, preprocessed, and unified EEG, iEEG, MEG datasets from OpenNeuro and Harvard EEG dataset into a single corpus for large-scale electrophysiology pretraining; spans tasks including visual perception, speech, and in-context learning.
  - **BEEGBench: Reproducible EEG, iEEG, MEG SFT dataset package:** Lightweight framework for config-centric building of SFT datasets.
  - **Publication:** Manuscript in preparation
- **Localizing One-shot Visual Perceptual Learning in Humans** New York, NY  
*Computational Neuroscience / Deep Learning* 2022 – 2024
  - **Top-down mechanism:** Proposed a neuro-inspired top-down mechanism in a deep neural network model that rivals human learning ability; analysis suggests perceptual-learning information concentrates in fusiform cortex.
  - **Publication:** First-author paper accepted at Nature Communications (preprint).
- **Routine EEG Foundation Model for Epilepsy Diagnosis** New York, NY  
*Self-supervised pretraining* 2024
  - **Electrode-agnostic representations:** Used BEEG and masked autoencoding to pretrain an electrode-agnostic model; evaluated transfer to epilepsy prediction on unseen subjects.
  - **Representation analysis:** Found epileptic representations exhibit higher intrinsic dimension than normal EEG, supporting increased signal complexity under pathology.
  - **Publication:** CNS 2025 poster, selected as oral presentation
- **Clinical NLP: Health-system Scale Language Models** New York, NY  
*Pretraining Infrastructure / Clinical Prediction* 2022 – 2023
  - **Research infrastructure:** Contributed debugging and infrastructure for pretraining large-scale clinical language models used for 30-day readmission prediction (*Nature*, 2023).
  - **Clinician-aligned sensitivity:** Quantified that classifier decisions align better with physician word sensitivity than bag-of-words/XGBoost baselines; poster at NeurIPS ML4H Workshop 2022.

- **Longitudinal Deep Learning for Metastatic Brain Cancer** New York, NY  
2022 – 2023  
*Medical Imaging Benchmarking*
  - **Benchmark contribution:** Developed a massive open benchmark showing longitudinal information improves performance on small tumors. Paper published in Nature communications.
- **Automated Detection of Cortical Spreading Depression (CSD) from EEG** New York, NY  
2021 – 2023  
*Graph Neural Networks + CNNs*
  - **Modeling:** Designed a novel architecture combining graph neural networks and convolutional neural networks to detect cortical spreading depressions from scalp EEG.
  - **Deployment-minded robustness:** Improves accuracy while enabling scalable deployment across different EEG sensor densities.
- **VoxelFormer (Mentored Project)** New York, NY  
2024 – 2025  
*fMRI Decoding / Parameter-efficient Modeling*
  - **Model:** Proposed a subject-independent architecture using voxel-level token merging for multi-subject visual decoding from fMRI; poster at IEEE/EMBS NER 2025.
- **ICD-9 Code Assignment (Mentored Project)** New York, NY  
2023  
*Clinical NLP / Imbalanced Learning*
  - **Method:** Investigated correlation priors between labels and showed large impacts on imbalanced clinical code prediction; presented at ACL Student Research Workshop 2023.

## OPEN SOURCE

---

- **x-Transformers (lucidrains/x-transformers)** Remote  
2024  
*Contributor*
  - **Bug fixes:** Contributed fixes related to positional encoding and masking in a widely-used Transformer implementation.

## COURSEWORK

---

- **Signal Processing:** Signal Processing; Image & Video Processing; Medical Imaging; Neural & Physiological Signal Processing
- **Stats & Probability:** Stochastic Processes; Detection and Estimation Theory
- **Optimization:** System Optimization Method
- **Psychology/Neuroscience:** Categories and Concepts; Sensation and Perception

## SKILLS

---

- **Programming Languages:** Python, C, MATLAB
- **ML / Software & Tools:** PyTorch, LaTeX
- **Domains:** EEG/iEEG/MEG, fMRI, clinical notes, medical imaging