WENJIE SUN

email: 12233339@mail.sustech.edu.cn; wj.sun@siat.ac.cn; wenjie1835@gmail.com Tel: +86 13895591107

Personal Website; Github; Google Scholar.

EDUCATION AND APPOINTMENTS

Northeastern University (NEU) **Shenyang, China**

Bachelor of Applied Chemistry 2018.09 – 2022.07

Southern University of Science and Technology (SUSTech) **Shenzhen, China**

Master of Electronic and Information Engineering 2022.09-2025.07

Mohamed bin Zayed University of Artificial Intelligence (MBZUAI) **Abu Dhabi, United Arab Emirates**

Research Engineer (Supervisor: Di Wang, Lijie Hu) 2025.12-Present

Major Courses: C++ Programming, MATLAB Programming, Modern Signal Processing, Mathematical Statistics, Linear Algebra, Principles of Automatic Control.

PUBLICATIONS

• **Wenjie Sun**, Lijie Hu, Di Wang. *"The Price of Amortized Inference in Sparse Autoencoders"* submitted to International Conference on Learning Representations (**ICLR 2026**). (Jun.2025 – Sep.2025) [Link]

• Jinning Yang, **Wenjie Sun**, Shi Wen. *"DiagECG: Discretized ECG Tokenization for Diagnostic Reasoning"* (**AAAI 2026**). (Dec.2024 – Jul.2025) [Link]

• **Wenjie Sun**, Zhile Yang, ..., Chengke Wu. *"Sparsification and Reconstruction from the perspective of Representation Geometry"* (**arXiv**). (Nov.2024 – May.2025) [Link]

• **Wenjie Sun**, Chengke Wu, ..., Zhile Yang. *"CauseTerML: Causal Learning via Term Mining for Assessing Review Discrepancies"* published in IEEE **Transactions on Artificial Intelligence**. (Jun.2023 – Nov.2024) [Link]

• **Wenjie Sun**, ..., Wenjiao Yao, Zhile Yang *"Fine-Tuning Enables State of Health Estimation in Lithium-Ion Batteries via Time Series Foundation Model"* published in **Energy**. (Nov.2023 – Dec.2024) [Link]

• **Wenjie Sun**, ..., Wenjiao Yao, Zhile Yang. *"State-of-charge estimation of sodium-ion batteries: A fusion deep learning approach"* published in **Journal of Energy Storage**. (Dec. 2022 - Mar. 2024) [Link]

RESEARCH INTERESTS

Causal Learning, Mechanistic Interpretability, Representation Learning.

RESEARCH EXPERIENCE

- **Mechanistic Interpretability**| Advisors: **Prof. Lijie Hu (MBZUAI), Prof. Di Wang (KAUST)** Nov.2024 – May.2025

Representative Work 1: The Price of Amortized Inference in Sparse Autoencoders

- We argue that there is a fundamental conflict between amortized inference-based Sparse Autoencoders (SAEs) and the instance-level optimality required for monosemantic features, and point out that the reconstruction-sparsity Pareto frontier metric pursued by current SAEs is not directly correlated with improvements in monosemanticity.
- To demonstrate the existence of this conflict, we analyze the training dynamics of SAEs under global reconstruction-sparsity constraints. We reveal an unreasonable trade-off relationship among various pathological phenomena (such as feature absorption, feature splitting, dead latents, etc.), indicating that these pathologies are not independent but rather a necessary compromise under the constraints of amortized inference.
- As a first step in this direction, we explore semi-amortized and non-amortized encoding methods. We demonstrate that they can significantly alleviate reconstruction errors and the dead latent problem, while also enhancing feature intervenability and downstream task performance, thereby providing a new direction for polysemy disentanglement.

Representative Work 2: Sparsification and Reconstruction from the perspective of Representation Geometry

- Proposed SAEManifoldAnalyzer (SAEMA), modeling SAE latent tensors on symmetric positive semi-definite matrix manifolds, revealing stratified representations in language model activations.

• Defined **local and global representations**, showing sparse encoding reduces feature overlap by merging semantics and adding dimensionality, elucidating monosemantic feature identification.
• Established **causal link** between local representation separability and reconstruction performance via geometric intervention, guiding SAE optimization under sparsity-reconstruction trade-offs.

- **Intersection of Causal Inference and NLP** | Advisors: **Prof.** Chengke Wu (Shenzhen Institutes of Advanced Technology, CAS) Jun. 2023 – Nov. 2024

Representative Work: Causal Learning via Term Mining for Assessing Review Discrepancies

• Proposed CauseTerML, a novel framework combining term mining and causal inference techniques to assess biases in innovation evaluation processes, particularly focusing on the impact of title technicality on patent authorization duration.
•Developed TerPOSMMI, a hybrid unsupervised term extraction method that integrates statistical and rule-based approaches, addressing challenges in handling interdisciplinary and time-sensitive terminology.
•Introduced Multi-character Mutual Information (MMI) to overcome limitations of traditional Point-wise Mutual Information, enhancing the ability to identify multi-word terms and normalize for word frequency effects.

- **Time series Forecasting (Battery state Estimation)** | Advisors: **Prof.** Zhile Yang, **Prof.** Wenjiao Yao (Shenzhen Institutes of Advanced Technology, CAS) Dec. 2022– Dec. 2024

Representative Work 1：Estimating Lithium Battery State of Health by Fine-tuning Time Series Base Model

• We fine-tuned TimeGPT, a time series foundation model, to achieve high-precision state of health estimation for lithium-ion batteries, eliminating the need for complex feature engineering traditionally required in battery health estimation.
•To enhance generalization, we utilized a diverse fine-tuning dataset comprising 143 lithium-ion batteries with six different cathode materials. The resulting model demonstrated strong performance on unknown battery domains, indicating its potential for rapid adaptation to novel scenarios.
•Our fine-tuned TimeGPT model outperformed the majority of state-of-the-art approaches, including supervised models, when compared against five categories of neural network-based models across 12 strong baselines.

Representative Work 2：State of Charge estimation of Sodium-Ion batteries based on fusion deep learning method

• Conducted comprehensive experimental studies on laboratory-fabricated sodium-ion batteries, providing valuable datasets for research on various cathode materials under different operating conditions.
•Proposed a novel fusion deep learning model combining BiLSTM and N-BEATS to address the unique challenges of SOC estimation in sodium-ion batteries, particularly the low voltage sensitivity to SOC.
•Developed an adaptive fusion algorithm to integrate BiLSTM and N-BEATS networks in parallel, demonstrating superior performance compared to individual models and offering a viable solution for battery management systems in sodium-ion batteries.

ACADEMIC THESIS

Undergraduate Graduation Thesis: NEU
Preparation and Electrocatalytic Performance Study of Polypyrrole-Derived Carbon Materials
Master's Graduation Thesis: SUSTech
Battery State Estimation of Sodium-ion Batteries based on Deep Learning Methods

HONORS AND AWARDS

National Scholarship (Highest honor, Top 0.3%) Sep.2023-Sep.2024
Outstanding graduate student in SUSTech. 2023. Jul.2022-Jul.2023
2022 CAS University Science and Art Fusion Creation Competition - Second Place Nov.2022
Northeastern University Chemistry Experiment Innovation Design Competition – Second Place Jun.2020
Northeastern University's Annual Outstanding Student Journalist Sep.2018-Sep.2019
Outstanding Volunteer for the Undergraduate Robotics Competition Oct.2018

EXTRACURRICULAR ACTIVITIES & SKILLS

Interests: Photography; Basketball; Travelling; Hiking; Fitness;
Content creator: https:// space.bilibili.com/416517790?spm_id_from=333.1007.0.0
Programming: C, C++, Python (Pytorch, Tensorflow, PyG), Mathematica, LATEX, R, MATLAB

English: TOEFL: 96, CET-6: 526, CET-4: 543.