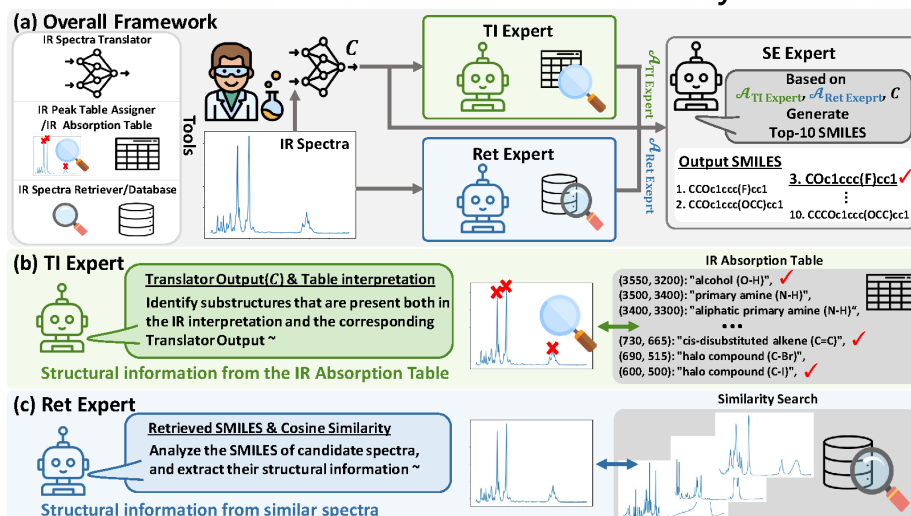


DATA SCIENCE & ARTIFICIAL INTELLIGENCE LAB (1)

In this Winter, we are mainly looking for students interested in AI for materials.

▪ Topic: AI for Materials

- Property prediction: crystal / composition representation learning, electronic structure (DOS)
- Synthesis planning: LLM-based precursor & synthesis condition prediction, retrosynthesis
- Inverse design: crystal generation, structure retrieval from measurements (e.g., XRD)
- Molecules & Chemistry: molecular representation learning, LLM agents for spectra (IR)
- Toward autonomous materials discovery with scientific LLM agents



LLM Agents for IR Spectra

MSP-LLM

To synthesize $\text{La}_{1.2}\text{Ca}_{1.8}\text{Mn}_2\text{O}_7$, what precursor set and synthesis operations should I use?

SE Expert Based on $\text{A}^2\text{TI}^2\text{Expert}$ & $\text{Ret}^2\text{Expert}$ Generate Top-10 SMILES

Output SMILES

1. CCOC1ccc(F)cc1
2. CCOC1ccc(OCC)cc1
3. CCOC1ccc(F)cc1
4. CCOC1ccc(OCC)cc1
5. CCOC1ccc(OCC)cc1
6. CCOC1ccc(OCC)cc1
7. CCOC1ccc(OCC)cc1
8. CCOC1ccc(OCC)cc1
9. CCOC1ccc(OCC)cc1
10. CCOC1ccc(OCC)cc1

SE Expert Based on $\text{A}^2\text{TI}^2\text{Expert}$ & $\text{Ret}^2\text{Expert}$ Generate Top-10 SMILES

SE Expert Based on $\text{A}^2\text{TI}^2\text{Expert}$ & $\text{Ret}^2\text{Expert}$ Generate Top-10 SMILES

SE Expert Based on $\text{A}^2\text{TI}^2\text{Expert}$ & $\text{Ret}^2\text{Expert}$ Generate Top-10 SMILES

SE Expert Based on $\text{A}^2\text{TI}^2\text{Expert}$ & $\text{Ret}^2\text{Expert}$ Generate Top-10 SMILES

SE Expert Based on $\text{A}^2\text{TI}^2\text{Expert}$ & $\text{Ret}^2\text{Expert}$ Generate Top-10 SMILES

SE Expert Based on $\text{A}^2\text{TI}^2\text{Expert}$ & $\text{Ret}^2\text{Expert}$ Generate Top-10 SMILES

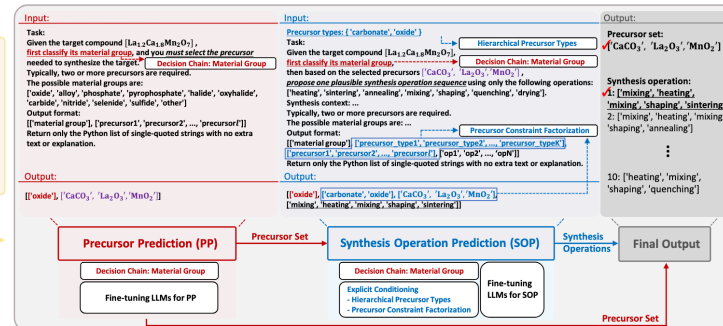
SE Expert Based on $\text{A}^2\text{TI}^2\text{Expert}$ & $\text{Ret}^2\text{Expert}$ Generate Top-10 SMILES

SE Expert Based on $\text{A}^2\text{TI}^2\text{Expert}$ & $\text{Ret}^2\text{Expert}$ Generate Top-10 SMILES

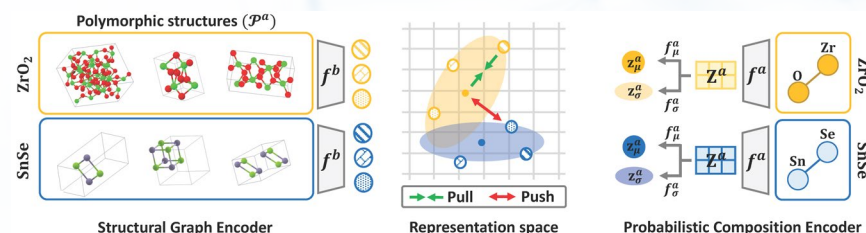
SE Expert Based on $\text{A}^2\text{TI}^2\text{Expert}$ & $\text{Ret}^2\text{Expert}$ Generate Top-10 SMILES

SE Expert Based on $\text{A}^2\text{TI}^2\text{Expert}$ & $\text{Ret}^2\text{Expert}$ Generate Top-10 SMILES

SE Expert Based on $\text{A}^2\text{TI}^2\text{Expert}$ & $\text{Ret}^2\text{Expert}$ Generate Top-10 SMILES



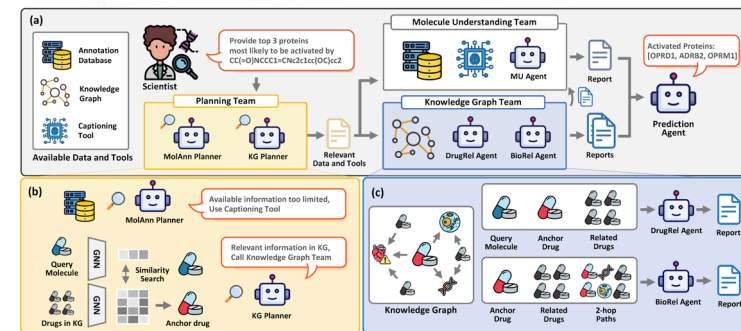
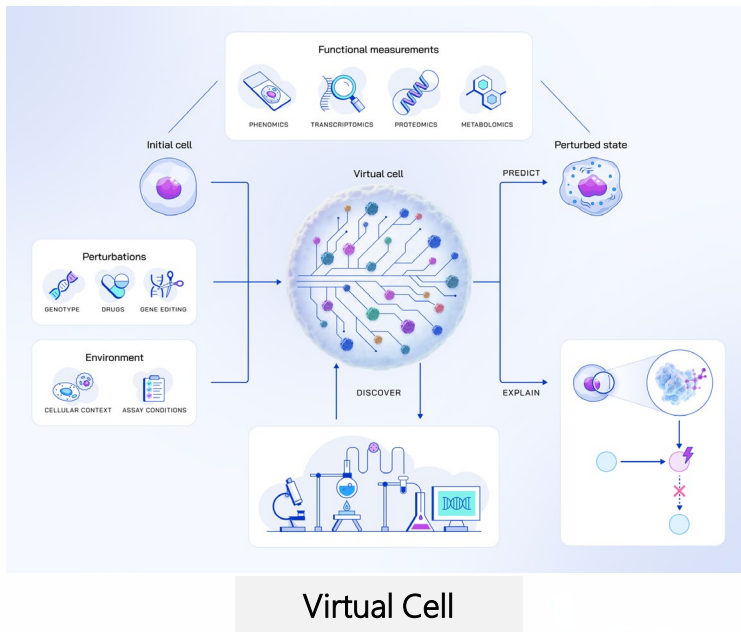
Materials Synthesis Planning



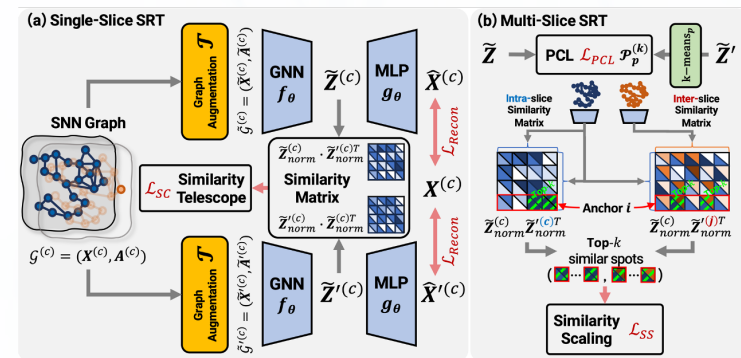
Our group's papers on these topics:

- <https://openreview.net/pdf?id=4UtSgrJ8N>
- <https://openreview.net/pdf?id=6UIOrnEUHL>
- <https://arxiv.org/pdf/2410.21341>
- <https://openreview.net/pdf?id=2lWh1G1W1l>
- <https://openreview.net/pdf?id=6bthH14pD8>

- Single-cell & spatial omics: cell representation learning, spatial transcriptomics
- Virtual cell: perturbation (transcriptional) response prediction, multi-agent reasoning
- Proteins & drug discovery: antibody-antigen structures, LLM agents for drug discovery
- Drug delivery: screening lipid nanoparticles (LNPs) via structure-ratio alignment



LLM Agents for Drug Discovery



Spatial Transcriptomics

Our group's papers on these topics:

<https://arxiv.org/pdf/2602.18885>

<https://arxiv.org/pdf/2506.15698>

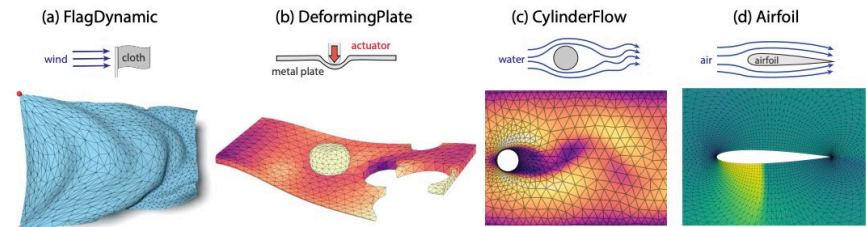
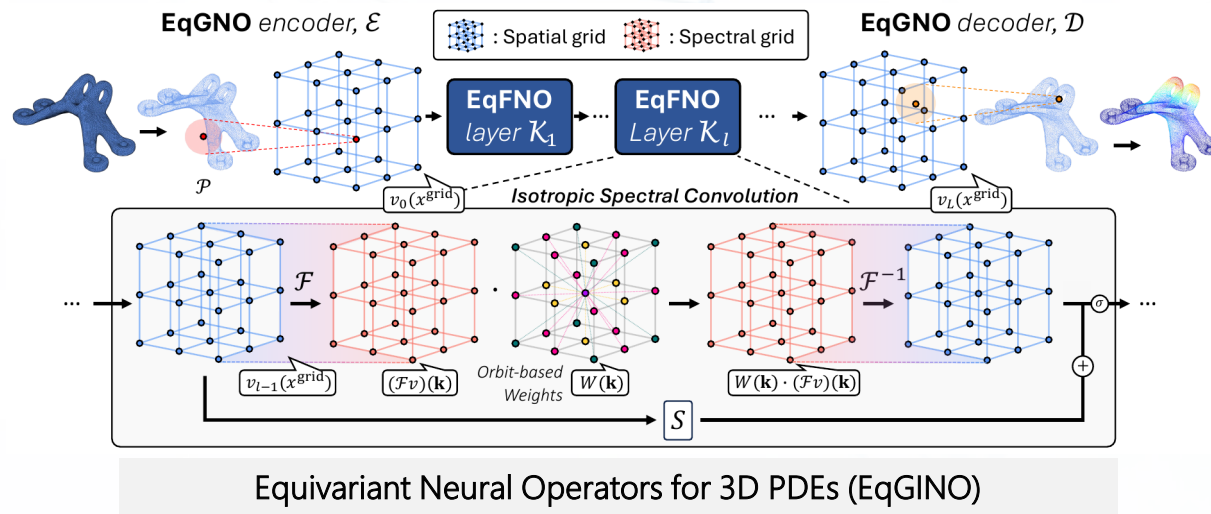
<https://arxiv.org/pdf/2502.17506>

<https://openreview.net/pdf?id=T4NJqDU2UI>

<https://www.biorxiv.org/content/10.64898/2026.07.08.737142v1>

DATA SCIENCE & ARTIFICIAL INTELLIGENCE LAB (3)

- **Topic:** AI for Engineering (Physical Systems)
 - Learning physical simulation: mesh-based GNN simulators for fluid dynamics
 - Neural operators for 3D PDEs: equivariant, geometry-informed, physics-embedded models
 - 3D geometry: E(3)-equivariant mesh neural networks for engineering structures



Our group's papers on these topics:

<https://arxiv.org/pdf/2606.03260>

<https://arxiv.org/pdf/2505.21572>

<https://arxiv.org/pdf/2511.12709>

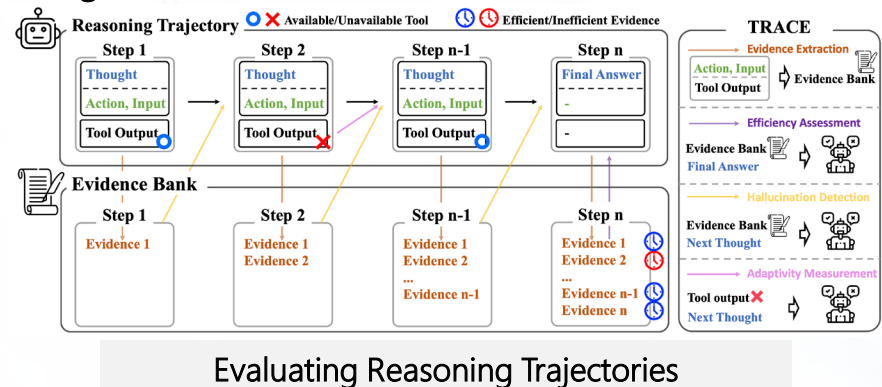
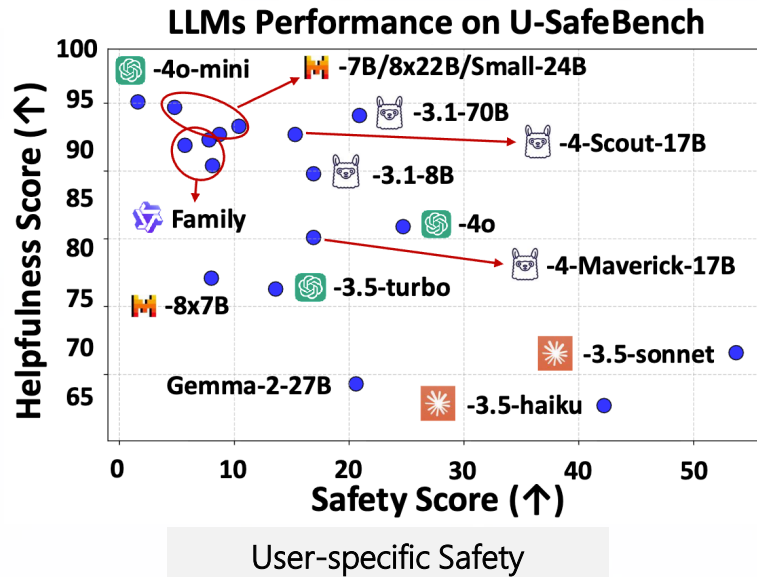
<https://www.sciencedirect.com/science/article/pii/S0043135426004550>

Mesh-based Simulation

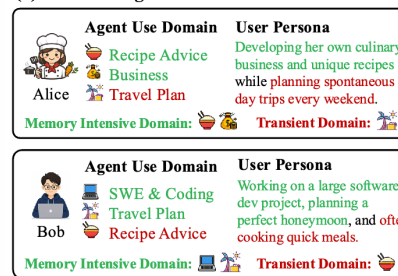
DATA SCIENCE & ARTIFICIAL INTELLIGENCE LAB (4)

▪ Topic: Agentic AI & Reasoning

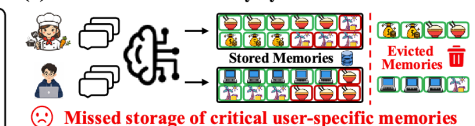
- Process-level evaluation: reasoning trajectories of tool-augmented agents, failure attribution
- Agent optimization: reward models for multi-turn agents, harness optimization from traces
- Memory & personalization: personalized long-term memory for long-horizon / embodied agents
- Safety & alignment: user-specific safety, safety alignment of reasoning models
- Efficient inference: speculative decoding, token pruning



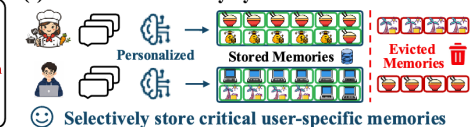
(a) Different Agent Use Patterns



(b) One-size-fits-all Memory System



(c) Personalized Memory System



Our group's papers on these topics:

<https://arxiv.org/pdf/2510.02837>

<https://arxiv.org/pdf/2605.17877>

<https://arxiv.org/pdf/2605.25535>

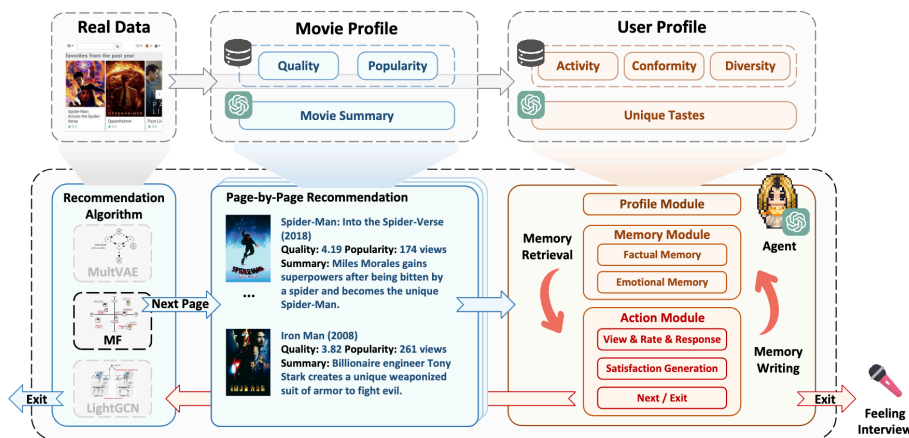
<https://arxiv.org/pdf/2608.23041>

<https://arxiv.org/pdf/2502.15086>

Personalized Agent Memory

DATA SCIENCE & ARTIFICIAL INTELLIGENCE LAB (5)

- **Topic:** Personalized Intelligence & Recommendation
 - LLM-based recommender systems: do LLMs understand collaborative / sequential signals?
 - LLM-based user simulators: building, evaluating, and diagnosing simulated users
 - Self-evolving recommender systems: LLM / coding agents that evolve models from feedback
 - Multi-modal & long-tail recommendation: missing modalities, continual user modeling



LLM-based User Simulator

Our group's papers on these topics:

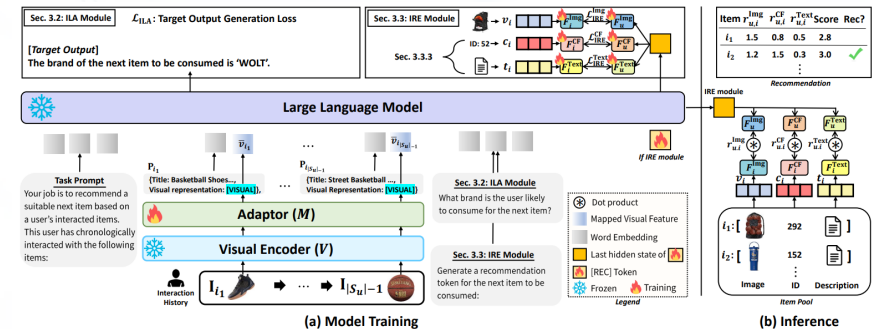
<https://openreview.net/pdf?id=fBLhmZ1aNN>

<https://openreview.net/pdf?id=vizM7B7vuW>

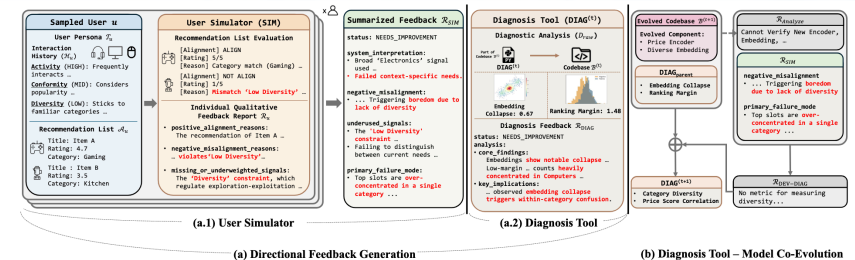
<https://arxiv.org/pdf/2502.13909>

<https://arxiv.org/pdf/2404.11343>

<https://arxiv.org/pdf/2609.25572>



Token-efficient LLM Recommender

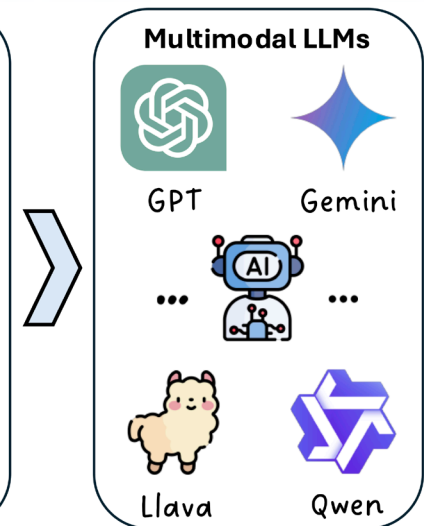
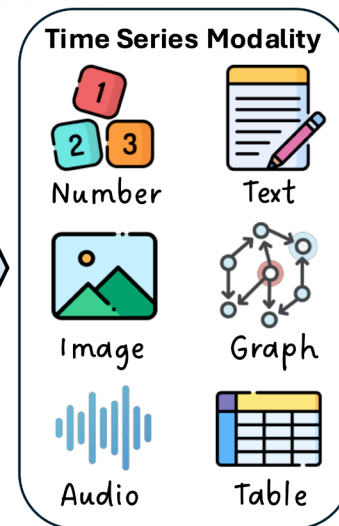
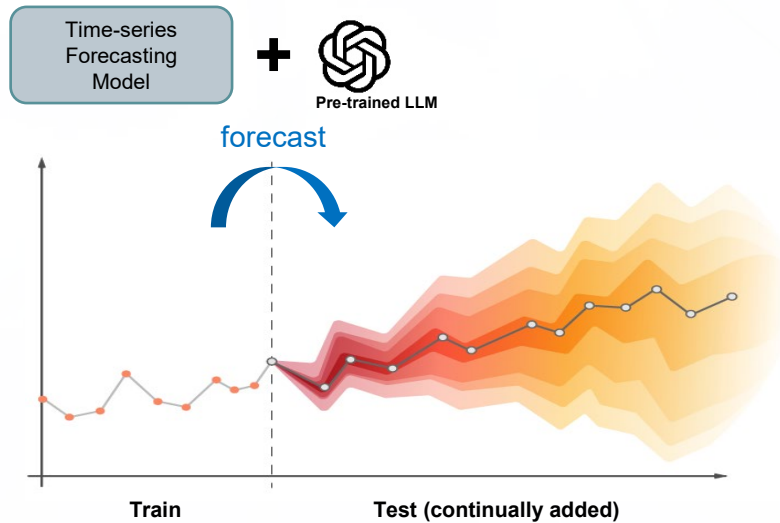


Self-Evolving Recommender System

DATA SCIENCE & ARTIFICIAL INTELLIGENCE LAB (6)

▪ Topic: Time-series Modeling

- Time-series foundation models: parameter-efficient fine-tuning to new domains
- Online forecasting under distribution shift: LLM prompting, continual forecasting
- Spatio-temporal modeling: traffic / sensor networks, self-explainable temporal GNNs
- Multi-modal time series & industrial applications (e.g., anomaly detection)



Multi-modal Time Series

Our group's papers on these topics:

<https://openreview.net/pdf?id=n8seTOinYs>

<https://openreview.net/pdf?id=QJGmolk1r9>

<https://arxiv.org/pdf/2406.03140>

TENTATIVE SCHEDULE (DEPENDS ON PROGRESS)

In this Winter, we are mainly looking for students interested in AI for materials.

- **Study & implement research papers published by our group (Week 1~4)**
 - Paper seminar / week (3 papers per week)
 - Checking progress and Q&A
 - Implement & reproduce each paper
 - 1:1 meeting available upon request
- **Participate in a project (Week 4~)**
 - Projects
 - 1. AI for Materials
 - 2. AI for Biology
 - 3. AI for Engineering (Physical Systems)
 - 4. Agentic AI & Reasoning
 - 5. Personalized Intelligence & Recommendation
 - 6. Time-series Modeling
 - ... many more
 - You can participate in the project you wish to participate (with graduate students)

DATA SCIENCE & ARTIFICIAL INTELLIGENCE LAB REQUIREMENT

- Requirement
 - Familiar with programming in Python
 - Knowledge in machine learning and deep learning
 - Experience in deep learning libraries such as PyTorch, Tensorflow, Keras is a strong plus
- Interns will team with graduate students
- If you have any topic you would like to work on, write about it in the application