

# Kaichen Xu

TEL: (+86) 18627184798 | E-mail: [kaichenxu358@gmail.com](mailto:kaichenxu358@gmail.com) | Homepage: <https://catchxu.github.io/>

## EDUCATION

Zhongnan University of Economics and Law (ZUEL)

09/2020 – 06/2024

**Major:** Bachelor of Science in Statistics; **GPA:** 92.21/100

**Course Highlight:** Python Data Analysis (95), Deep Learning (93), Artificial Intelligence (93), Optimization Theory (91), Principles of Database System (93)

## RESEARCH INTERESTS

Generative Model; Computational Biology and Bioinformatics; Theory of Computation and Deep Learning

## PUBLICATIONS

†: equal contribution

- [1] **Kaichen Xu**, Yan Lu, and et al. Detecting and Dissecting Anomalous Anatomic Regions in Spatial Transcriptomics with STANDS. *Nature Communications*, 2024.
- [2] **Kaichen Xu**<sup>†</sup>, Yueyang Ding<sup>†</sup>, and et al. Domain Adaptive and Fine-grained Anomaly Detection for Single-cell Sequencing Data and Beyond. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 2024.
- [3] **Kaichen Xu**, Pengyu Xie, and et al. Large-scale Gene Topic Modeling for Single-cell Multi-omics Retrieval Augmented Generation. *Nature Methods*, 2024. Under Review.
- [4] **Kaichen Xu**, Yan Lu. Align Anything: Learning Coherent Representation with Multi-domain Independence Discrepancy. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024. Under Review.
- [5] **Kaichen Xu**<sup>†</sup>, Qilong Wu<sup>†</sup>, and et al. Multimodal Anomalous Tissue Region Detection Enhanced with Spatial Transcriptomics. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2024. Under Review.
- [6] **Kaichen Xu**, Yan Lu. Series Anomaly Inference with Nonlinear Markov to Circumvent Walking a Tightrope. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2024. Under Review.
- [7] **Kaichen Xu**<sup>†</sup>, Kainan Liu<sup>†</sup>, and et al. An Integrated Approach for Detecting and Subtyping Anomalous Single Cells across Samples and Modalities. *Genome Biology*, 2024. Under Review.
- [8] **Kaichen Xu**. Estimation and Application of the Convergence Bounds for Nonlinear Markov Chains. *arXiv preprint arXiv:2209.12834*, 2022.

Other publications: 4 co-authored (non-first) papers; 2 granted patents

## RESEARCH EXPERIENCES

**Large-scale Gene Topic Modeling for Single-cell Sequencing Data Retrieval-Augmented Generation** 07/2024 – Present

- Developed the first Large Language Model (LLM) designed for Retrieval-Augmented Generation of single-cell datasets. This model not only can analyze single-cell datasets to deliver nuanced pathological and biological interpretations but also of automatically retrieve and integrate the most relevant datasets from extensive databases based on user requirements.
- Employed the large-scale pretrained topic models in gene embedding space to extract data-driven gene interactions; utilized Gene Regulatory Networks and the Gene Ontology database to extract knowledge-driven gene interactions. These gene interactions are structured into a knowledge graph prior, which is then integrated into LLM.
- Under review by Nature Methods.

**Align Anything: Learning Coherent Representation with Multi-domain Independence Discrepancy** 04/2024 – 11/2024

- Focused on the unified challenges in multi-modal, view, and domain datasets fusion and representation learning. Proposed a general loss, Multi-domain Independence Discrepancy (MID), for different types of data alignment and representation learning.
- Theoretically demonstrated that MID loss has excellent statistical and optimization properties (e.g., unbiasedness, robustness, and differentiability). Experimentally demonstrated that substituting or augmenting the original loss with MID loss led to improved performance in most widely used models across their respective tasks.
- Under review by the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

## **Detecting and Dissecting Anomalous Anatomic Regions in Spatial Transcriptomics with STANDS** 10/2023 – 08/2024

- Explored three interrelated challenges of disease heterogeneity analysis on Spatial Transcriptomics. Revealed the population-level and individual-specific pathogenic factors significantly affected the discovery of cancer microenvironment.
- Proposed STANDS, a novel framework based on multi-modal Generative Adversarial Networks. Achieved the state-of-the-art model in the identification of pathological anatomic tissues by fusing spatial gene expression data with H&E images.
- Successfully published research at Nature Communications as an undergraduates.

## **Domain Adaptive and Fine-grained Anomaly Detection for Single-cell Sequencing Data and Beyond** 09/2023 – 05/2024

- Discovered and mathematically theorized that domain shifts prevalent in multi-sample and multi-domain datasets can impact the accuracy of anomaly detection. Also provided new insights into the limitations of existing anomaly detection methods.
- Inspired by the theories about Maximum Mean Discrepancy and Ramanujan's master theorem, proposed an innovative framework, ACSleuth, for discovering anomalous cell types and subtypes in single-cell sequencing data. Extensive experiments and theoretical analysis demonstrate its superiority over existing methods and robustness to domain shifts.
- Successfully presented finds at the International Joint Conference on Artificial Intelligence as an undergraduates.

## **LEADERSHIP EXPERIENCES**

### **Organizer and Speaker of Internet Academic Seminar Group about Machine Learning** 09/2023 – Present

- Led weekly online seminars on machine learning technologies, engaging over 30 participants from various universities.
- Delivered a presentation once a month on the latest technical papers, sharing personal insights and innovative ideas. Inspired by them, the participants published over 10 conference papers in the field of computer science or computational biology.

### **Research Assistant and Mentor at Sunlab (PI: Xiaobo Sun)** 03/2023 – Present

- Managed the lab's servers and computing resources, ensuring high reliability, and efficient access for the entire research team.
- Conducted over 30 one-on-one interviews with lab candidates. Provided tailored mentorship and guidance to newly joined students, leading to a sophomore successfully publishing a paper at ACM SIGKDD.

### **Teaching Assistant for Machine Learning (Master) and Optimization Theory (Bachelor)** 09/2024 – 01/2025

- Provided personalized academic consulting to over 20 graduates and 50 undergraduates, effectively guiding them through complex concepts and resolving learning challenges.
- Taught specialized section on language models in computational biology, based on personal research experience.

### **Technical Leader at Doctor Yu, Large Scale Intelligent Nucleic Acid Test Management System** 02/2022 – 12/2022

- Designed and implemented data-driven technical solutions for optimizing nucleic acid testing appointment scheduling using reinforcement learning, addressing challenges such as unpredictable arrival times and low allocation efficiency for subscribers.
- Successfully deployed the system across over 10 sites, including hospitals and campuses, managing an average of 10,000 tests per day. The system received widespread praise from local residents and reported by television for its impact and effectiveness.

## **ACADEMIC PRESENTATIONS**

---

Oral Presentation at the International Joint Conference on Artificial Intelligence (IJCAI)	08/2024
Oral Academic Seminar at Dept. of Comp. Sci. & Dept. of Biomed. Info., Emory University	10/2023
Oral Academic Presentation at Dept. of Comp. Sci., Indiana University	07/2023

## **AWARDS & HONORS**

---

National Collegiate Market Research and Analysis Competition Third Prize (Top 2% in China)	04/2023
Undergraduate Comprehensive Commendation for "Research Star" (Top 1 in ZUEL)	12/2022
Contemporary Undergraduate Mathematical Contest in Modeling First Prize (Top 2% in Hubei province)	10/2022
Kangteng National Collegiate Business Case Analysis Competition Third Prize (Top 0.6% in China)	06/2022
Zhongnan University of Economics and Law Mathematics Competitions First Prize (Top 1 in ZUEL)	12/2021

## **TECHNICAL SKILLS**

---

**Programming Languages:** Python (PyTorch, TensorFlow, Pandas, NumPy, Scanpy, AnnData, etc.), R, C/C++, PHP  
**Technologies and Tools:** Latex, Linux, GitHub, Docker, MATLAB, MySQL, Markdown, HTML/CSS