Ziyang Song

Stocker Center 313, School of Electrical Engineering and Computer Science, Ohio University

ziyangs@ohio.edu Website LinkedIn Google scholar

Research Summary:

My research focuses on **AI for health**, with a focus on generative AI and statistical machine learning. My vision is to build **trustworthy LLMs** to address biomedical challenges.

1. Trustworthy LLM in healthcare: I design novel algorithms to enhance explainability, uncertainty analysis, and prediction safety for biomedical LLMs.
2. Generative AI for health time series: I focus on time-series representation learning for biosignals, clinical measurements, and longitudinal medical records.
3. Statistical machine learning for medical NLP: I design probabilistic models and statistical inference for interpretable medical representations using electronic health records.

Full-time Professional Experience:

- **Assistant Professor** 08/2025 – present
  School of Electrical Engineering and Computer Science, Ohio University, Athens, Ohio, United States

Education:

- **Ph.D. Computer Science** 09/2020 – 05/2025
  School of Computer Science, McGill University, Montreal, Canada
  Affiliation: Mila - Quebec AI Institute
  Advisor: Yue Li
- **M.Sc. Quality Systems Engineering** 09/2018 – 11/2019
  Concordia Institute for Information Systems Engineering, Concordia University, Montreal, Canada
  Advisor: Nizar Bouguila
- **B.Eng. Computer Science** 09/2014 – 06/2018
  University of Shanghai for Science and Technology, Shanghai, China

Publications:

Conference Papers:

1. ***Song, Z.***, Lu, Q., Xu, Q., Buckeridge, DL, Li, Y. (2024). TimelyGPT: Extrapolatable Transformer Pre-training for Long-term Time-Series Forecasting in Healthcare. *ACM BCB*, **Rising Star Award**.
2. Wang, Z., Wang, R., ***Song, Z.***, Buckeridge, DL, Li, Y. (2024). MixEHR-Nest: Identifying subphenotypes within electronic health records through hierarchical guided-topic modeling. *ACM BCB*.
3. ***Song, Z.***, Lu, Q., Xu, He, Z., Buckeridge, DL., Li, Y. (2024). Bidirectional generative pre-training for improving healthcare time-series representation *learning. MLHC (PMLR).*
4. ***Song, Z.***, Hu, Y., Verma, A., Buckeridge, DL., Li, Y. (2022). Automatic phenotyping by a seed-guided topic model. *KDD,* **HealthDay Best Paper award**.
5. ***Song, Z.***, Toral, XS., Xu, Y., Liu, A., Guo, L., Powell, G., Verma, A., Buckeridge, DL., Marelli, A., Li, Y. (2021). Supervised multi-specialist topic model with applications on large-scale electronic health record data. *ACM BCB*.
6. ***Song, Z.***, Bregu, O., Ali, S., Bouguila, N. (2019). Variational inference of finite asymmetric gaussian mixture models. *IEEE SSCI*.
7. ***Song, Z.***, Ali, S., Bouguila, N. (2019). Bayesian learning of infinite asymmetric gaussian mixture models for background subtraction. *ICIAR*.

Journal Papers:

1. ***Song, Z.***, Lu, Q., He, Z., Buckeridge, DL., Li, Y. (2025). TrajGPT: Healthcare Time-Series Representation Learning for Trajectory Prediction. *IEEE Journal of Biomedical and Health Informatics*. Accepted, to appear.
2. ***Song, Z.***, Lu, Q., Xu, Q., Buckeridge, DL, Li, Y. (2025). TimelyGPT: Extrapolatable Transformer Pre-training for Long-term Time-Series Forecasting in Healthcare. *Health Information Science and Systems.* Accepted, to appear.
3. Zou, Y., Pesaranghader, A., ***Song, Z.***, Verma, A., Buckeridge, DL., Li, Y. (2022). Modeling electronic health record data using an end-to-end knowledge-graph-informed topic model. *Scientific Reports*.
4. ***Song, Z.***, Ali, S., Bouguila, N. (2021). Bayesian inference for infinite asymmetric gaussian mixture with feature selection. *Soft Computing*.
5. ***Song, Z.***, Ali, S., Bouguila, N. (2020). Background subtraction using infinite asymmetric Gaussian mixture models with simultaneous feature selection. *IET Image Processing*.
6. ***Song, Z.***, Ali, S., Bouguila, N. Fan, W. (2020). Nonparametric hierarchical mixture models based on asymmetric gaussian distribution. *Digital Signal Processing*.

Preprints:

1. Shen, J., He, Z., ***Song, Z.***. (2025). SMIs: Semantic Medical IDs with Medical Ontology. Submitted to *AAAI SECURE-AI4H Workshop*.
2. ***Song, Z.***, Yang, Z., Zabad, S., MA Legault, MA., Li, Y. (2025). PheCode-guided multi-modal topic modeling of electronic health records improves disease incidence prediction and GWAS discovery from UK Biobank. ISMB GLBIO. Abstract for oral presentation. Submitted to *Briefs in Bioinformatics.*
3. ***Song, Z.***, Xu, M., Latour, F., Gravel, S., Ho, V., Lettre, G., Li, Y. (2024). AI-driven approach for computational phenotyping with CARTaGENE cohort. Scientific Meeting of the Canadian Translational Geroscience Network. Abstract for oral presentation.

Presentations:

Oral Presentation:

1. ***Song, Z.*** et al. (2024, Nov 22-25). TimelyGPT: Extrapolatable Transformer Pre-training for Long-term Time-Series Forecasting in Healthcare. *ACM BCB*. Shenzhen, China.
2. ***Song, Z.*** et al. (2024, Sep 5-6). AI-driven approach for computational phenotyping with CARTaGENE cohort. *Scientific Meeting of the Canadian Translational Geroscience Network*. Montreal, Canada.
3. ***Song, Z.*** et al. (2024, Apr 8-13). Tutorial - phenotyping and PheWAS using MixEHR-seed. *Tokyo Symposium on Genomic Medicine, Therapeutics and Health*. Tokyo, Japan.
4. ***Song, Z.*** et al. (2023, May 15-18). MixEHR-SAGE: A seed-guided topic model to improve phenotyping for PheWAS analysis in UK Biobank data. *ISMB GLBIO*. Montreal, Canada.
5. ***Song, Z.*** et al. (2022, Aug 14-18). Automatic phenotyping by a seed-guided topic model. *KDD*. Washington DC, U.S.
6. ***Song, Z.*** et al. (2021, Aug 1-4). Supervised multi-specialist topic model with applications on large-scale electronic health record data. *ACM BCB*. Virtual.

Poster Presentation:

1. ***Song, Z.*** et al. (2024, Dec 15). TrajGPT: Healthcare Time-Series Representation Learning for Trajectory Prediction. *NeurIPS 2024 Workshop TSALM*. Vancouver, Canada
2. ***Song, Z.*** et al. (2024, Aug 16-17). Bidirectional generative pre-training for improving healthcare time-series representation learning. *MLHC*. Toronto, Canada
3. ***Song, Z.*** et al. (2022, Oct 21). MixEHR-Seed: automatic phenotyping by a seed-guided topic model. *50th Anniversary of SOCS at McGill University*. Montreal, Canada.
4. ***Song, Z.*** et al. (2019, Aug 27-29). Bayesian learning of infinite asymmetric gaussian mixture models for background subtraction. *ICIAR*. Waterloo, Canada

Honors and Awards:

Academic Honors:

- ***FRQNT Provincial Doctoral Research Scholarship*** 05/2023 – 09/2025
- McGill Research Stipend 09/2020 – 04/2024
- Grad Excellence Award in Computer Science 09/2020 – 04/2024
- Faculty of Science Grad Supplement Award 09/2022 – 08/2024
- SOCS Grad Stimulus Initiative Award 09/2020 – 08/2022
- Jackie Cheung Graduate Award 09/2020 – 08/2021
- Concordia Master Research Stipend 09/2018 – 10/2019

Awards:

- ACM BCB Rising Star Award 2024
- ***KDD Healthy Day Best Paper Award*** 2022

Internship Experience:

- **Research Intern –** Centre University Hospital (CHU) Sainte–Justine 06/2023 – 12/2024
  - Developed a clinical decision-support system using AI-driven models to track patient diagnoses, monitor disease progression, and analyze cancer registration.
  - Applied a probabilistic AI model (MixEHR-SAGE) to infer expert-guided phenotype distributions for 50K individuals in the CARTaGENE cohort. Presented in oral presentation.
- **International Research Intern -** Nanyang Technology University 01/2020 – 06/2020
  - Developed probabilistic models using stochastic optimization techniques to analyze retail data, optimizing product recommendations and promotions tailored to patient profiles.
  - Led to a significant improvement for item recommendations and patient-targeted promotions.
- **Data Analyst Intern -** Shanghai MetaLab 07/2017 – 12/2017
  - Developed a click-model for a recommender system and implemented machine learning algorithms to analyze patent data. Supported data analysis and visualization to aid in marketing decisions.
  - Developed a web crawler, database system, and back-end data interfaces to process around 10 million public patents. Maintained the database for efficient and reliable data analysis.
- **Data Analyst Intern -** Shanghai Qingyue Environment Protection Center (NGO) 01/2017 – 03/2017
  - Developed an environmental data platform that provides public access to weather data, aimed at supporting efforts to reduce environmental pollution in China.
  - Developed a web crawler, database system, and back-end data interfaces to efficiently process and manage publicly available weather datasets.

Teaching Experience

Ohio University

- **Instructor** Fall 2025
  - AI 5010/4010 Foundations of Deep Learning

McGill University

- **Guest Lecturer** Fall 2024
  - COMP 565 Machine Learning in Genomics and Healthcare
- **Guest Lecturer** Fall 2023
  - COMP 565 Machine Learning in Genomics and Healthcare
- **Teaching Assistant** Winter 2023
  - COMP 551 Applied Machine Learning
- **Teaching Assistant** Fall 2022
  - COMP 565 Applied Machine Learning

Mentor Experience:

Ohio University (2025/07 - present)

- UW–Madison master student Jiahe Shen 07/2025 – present
  - Mentored in applying semantic IDs in biomedical LLMs. Submitted to AAAI workshop 2026.
- Northwestern master student Zihao Li 07/2025 – present
  - Mentored in applying representation learning of medical concepts in hyperbolic spaces.
- Brown master student Yiran Ding 07/2025 – present
  - Mentored in applying constative learning and optimal transport in biomedical LLMs.
- OSU bachelor student Xiaohang Xie 07/2025 – present
  - Mentored in exploring confidence calibration in clinical reasoning and LLMs.
- OSU bachelor student Yang Hu 07/2025 – present
  - Mentored in exploring prediction confidence and rejection in LLMs.
- OSU bachelor student Zhongning Deng 07/2025 – present
  - Mentored in exploring uncertainty quantification and for LLM outputs.

McGill University (2020/09 - 2025/05)

- McGill master student Ruilin Wang 09/2024 – 2025/05
  - Mentored in applying LLMs in healthcare time-series data such as biosignals.
- McGill master student Bo Hong Wang 09/2024 – 2025/05
  - Mentored in applying LLMs in healthcare time-series data such as lab tests in ICU data.
- McGill undergraduate and master student Ziqi Yang 09/2022 – 2025/05
  - Mentored in applying MixEHR-Seed on UKB dataset, presented to ISMB GLBIO 2023.
  - Placement: master student, McGill University
- McGill undergraduate and master student Ruohan Wang 09/2022 – 09/2024
  - Mentored in designing MixEHR-Nest on Quebec's PopHR dataset, published at ACM BCB 2024.
  - Placement: Ph.D. student, Brown University
- McGill undergraduate student Hao Xu 05/2023 – 09/2023
  - Mentored in applying TimelyGPT on biosignals, published at ACM BCB 2024.
  - Placement: Ph.D. student, University of Pennsylvania
- McGill undergraduate student Ziyu Zhao 05/2023 – 09/2023
  - Mentored in applying TimelyGPT on Quebec PopHR database.
  - Placement: master student, McGill University
- McGill undergraduate student Yuanyi Hu 05/2021 – 05/2022
  - Mentored in designing MixEHR-Seed, published at KDD 2022.
  - Placement: master student, Columbia University
- McGill undergraduate student Yixin Xu 09/2020 – 09/2021
  - Mentored in applying MixEHR-S, published at ACM BCB 2021.
  - Placement: master student, Duke University

Services:

Conference Reviewer:

- ACM SIGSPATIAL (2025)
- KDD (2025, 2024)
- ICML Workshop FMSD (2025)
- IJCAI (2025)
- ICLR (2026, 2025)
- NeurIPS Workshop TSALM (2024)
- ML4H (2024)

Journal Reviewer:

- Applied Intelligence (2025)

- Big Data Mining and Analytics (2025, 2024)

Ohio University:

- Member of PhD student committee (2025-present)
- Member of CS/AI Assessment & Accreditation (2025-2026)

McGill University and Mila:

- Evaluation committee for Mila's Supervision Request Process (2024)

Referees:

- Dr. Yue Li. Associate Professor in School of Computer Science, McGill University. send.Li.D38D6CE548@interfoliodossier.com
- Dr. David Buckeridge. Professor in School of Population and Global Health, McGill University. send.Buckeridge.BC61C28AEC@interfoliodossier.com
- Dr. Archer Yi Yang. Associate Professor of Department of Mathematics and Statistics, McGill University. send.Yang.BCEBBEFEC2@interfoliodossier.com