Ziyang Song

School of Computer Science, McGill University, Montreal, Canada

ziyang.song@mail.mcgill.ca Website LinkedIn Google scholar

I am a PhD candidate specializing **AI for health**, with a focus on **generative modelling and probabilistic modelling**. In **generative modelling**, I focus on time-series representation learning for biosingals, clinical measurements, and longitudinal medical records, enabling health trajectory analysis and various clinical tasks. In **probabilistic modelling**, my work models interpretable medical representations and supports clinical decision-making on electronic health records. My expertise includes:

- **Deep Learning:** Transformer and foundation models, SSM, RNN, Neural ODE
- **Deep Generative Models:** VAE, Normalizing Flows, deep latent variable models
- **Probabilistic learning:** latent variable models, topic models, mixture models, statistical inference

Education:

- **Ph.D. Computer Science** 09/2020 – Expected 04/2025
  School of Computer Science, McGill University, Montreal, Canada
  Advisor: Yue Li
  Thesis: Probabilistic and generative models for electronic health records
- **M.Sc. Quality Systems Engineering** 09/2018 – 10/2019
  Concordia Institute for Information Systems Engineering, Concordia University, Montreal, Canada
  Advisor: Nizar Bouguila
  Thesis: Nonparametric Bayesian models based on asymmetric Gaussian distributions
- **B.Eng. Computer Science** 09/2014 – 06/2018
  University of Shanghai for Science and Technology, Shanghai, China

Research Experience:

- **PhD Research Assistant -** McGill University 03/2019 – Present
  - **Generative AI:**
    - Developed a foundation model (BiTimelyGPT) using a novel bidirectional generative pre-training on biosignals and longitudinal EHR data. Published at MLHC 2024 (PMLR).
    - Developed a foundation model (TimelyGPT) for forecasting patient healthcare trajectories using biosignals and longitudinal EHR data. Published at ACM BCB 2024 with oral presentation.
    - Explored the applications of LLMs on time series and tabular clinical data.
  - **Probabilistic AI:**
    - Developed a multi-modal, expert-guided topic model (MixEHR-Seed) for medical records. Published at KDD 2022 and won **Healthday Best Paper Award**.
    - Developed a multi-modal, self-supervised topic model (MixEHR-S) for medical records. Published at ACM BCB 2021.
- **Master Research Assistant -** Concordia University 03/2019 – 12/2019
  - Conducted research on probabilistic graphical models and statistical inference approaches for anomaly detection and quality analysis in the CV domain.

Selected 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, Oral presentation.
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, H., Zhu, Buckeridge, DL., Li, Y. (2024). Bidirectional generative pre-training for improving healthcare time-series representation learning. Machine Learning for Healthcare (MLHC) and Proceeding of Machine Learning Research (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. 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.
2. ***Song, Z.***, Ali, S., Bouguila, N. (2021). Bayesian inference for infinite asymmetric gaussian mixture with feature selection. Soft Computing.
3. ***Song, Z.***, Ali, S., Bouguila, N. (2020). Background subtraction using infinite asymmetric Gaussian mixture models with simultaneous feature selection. IET Image Processing.
4. ***Song, Z.***, Ali, S., Bouguila, N. Fan, W. (2020). Nonparametric hierarchical mixture models based on asymmetric gaussian distribution. Digital Signal Processing.

Preprints:

1. ***Song, Z.***, Lu, Q., Zhu, M., Buckeridge, DL., Li, Y. (2024). TrajGPT: Healthcare Time-Series Representation Learning for Trajectory Prediction. NeurIPS 2024 Workshop Time Series in the Age of Large Models (TSALM). Submitted to ICLR, under review.
2. ***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. BMC Medical Informatics and Decision Making, under review.
3. ***Song, Z.***, Yang, Z., Wang, R., Zabad, S., MA Legault, MA., Li, Y. (2023). MixEHR-SAGE: A seed-guided topic model to improve phenotyping for PheWAS analysis in UK Biobank data. ISMB GLBIO. Abstract for oral presentation. Journal of Biomedical Informatics, under review.

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). Practical - 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 (HealthDay Best Paper award). 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 Doctoral Research Scholarship*** 05/2023 – Present
- McGill Research Stipend 09/2020 – Present
- Grad Excellence Award in Computer Science 09/2020 – Present
- 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:

- KDD healthy day best paper award 09/2020 – 08/2021

Professional Experience:

- **Research Intern -** Centre de recherche du CHU Sainte–Justine 06/2023 – Expected 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:

- **Guest Lecturer -** McGill University Fall 2024
  - Course: COMP565 Machine Learning in Genomics and Healthcare
  - Title: Time-series Transformer for representation learning and health-related forecasting
- **Guest Lecturer -** McGill University Fall 2023
  - Course: COMP565 Machine Learning in Genomics and Healthcare

- Title: Time-series Transformer in EHR
- **Teaching Assistant** – McGill University Winter 2023
  - Course: COMP551 Applied Machine Learning
  - Duties: Office hours, tutorials, grading, design and evaluate final project.
- **Teaching Assistant** – McGill University Fall 2022
  - Course: COMP551 Applied Machine Learning
  - Duties: Office hours, tutorials, grading, research project instructor.

Mentor Experience:

- Master student Ruilin Wang 09/2024 – Present
  - Mentored in applying LLMs in healthcare time-series data such as biosignals.
- Master student Bo Hong Wang 09/2024 – Present
  - Mentored in applying LLMs in healthcare time-series data such as lab tests in ICU data.
- Undergraduate and master student Ziqi Yang 09/2022 – Present
  - Mentored in applying MixEHR-Seed on UKB dataset, presented to ISMB GLBIO 2023.
  - Placement: master student, McGill University
- Undergraduate and master student Ruohan Wang 09/2022 – 09/2024
  - Mentored in applying MixEHR-Seed on MIMIC dataset, presented to ISMB GLBIO 2023.
  - Mentored in designing MixEHR-Nest on Quebec's PopHR dataset, published at ACM BCB 2024.
  - Placement: master student, McGill University
- 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
- Undergraduate student Ziyu Zhao 05/2023 – 09/2023
  - Mentored in applying TimelyGPT on Quebec PopHR database.
  - Placement: master student, McGill University
- Undergraduate student Yuanyi Hu 05/2021 – 05/2022
  - Mentored in designing MixEHR-Seed, published at KDD 2022.
  - Placement: master student, Columbia University
- 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:

- KDD (2024), NeurIPS Workshop TSALM (2024), ML4H (2024), ICLR (2025)

Journal Reviewer:

- Big Data Mining and Analytics (2024)

McGill University and Mila:

- Evaluation committee for Mila's Supervision Request Process (2024)
  - Assist professors in reviewing the applications of MSc and PhD students

Referees:

- Dr. Yue Li. Assistant 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