# HONGGEN ZHANG

+1808-206-5018 honggen@hawaii.edu https://honggen-zhang.github.io

I have experience in developing AI methods to address biomedical research and clinical applications such as large language model-based mRNA embedding and AI-based ECG data analysis. I am currently also working on the microRNA to find the target site of a particular mRNA which is interpreted by the cross-attention score from the transformer-based model. This method utilized a large training dataset and included different data modalities. I am also interested in modeling the dynamic system of biological processes such gene expression process.

## Education

**University of Hawaii at Manoa** 08/2019 – 05/2025
Computer and Electrical Engineering -System and Data Science Ph.D HI, USA
**AI Institute in Dynamic system** Affiliate graduate student 08/2021 – 05/2025
**Harvard University** Visiting student 11/2024 – 02/2025 MA, USA
**North China University of Technology** 09/2016 – 05/2019
Applied Mathematical M.S Beijing, China

## Research Projects – NLP

**1. Alignment Large Language Model(LLM) - Project Website link** 11/2023 –

- Large-langue models (LLMs) provide the fundamentals to achieve artificial general intelligence (AGI). However, the fine-tuning process is very expensive. I use response embedding to select **high-quality data** for efficient and cheaper alignment of LLMs. We **save up to 60% of annotators' work**.

**2. Contrastive knowledge graph embedding without bias - Project Website link.** 08/2022 – 02/2024

- I have developed a knowledge graph representation learning model that considers the **bias in data sampling**. The method considers both text and graph modalities to remove the sampling bias, thus achieving the **SOTA result** on the link prediction.

**3. How news evolves - Project Website, Project Website link.** 01/2020 – 08/2022

- To distinguish the different types of news Medea, I proposed an **mathematical model** to characterize the of **news evolution** in term of the graph dynamic changing.

## Research Projects –AI4Science

**Nonlinear Dynamic System Representation Methods** 03/2023 –

- For the complex high-dimension dynamic system, it is hard to find the governing equation to describe it. Thus, I propose a **data-driven** and machine learning-based linear representation method for **nonlinear dynamic systems**. This method can stability model the long-term prediction of nonlinear dynamic systems.

**mRNA embedding - Project Website** 5/2024 –

- mRNA sequence has a similar feature to the natural language sequence. To achieve a better downstream task, such as mRNA translation efficiency, mRNA representation is necessary. A **contextual-based mRNA embedding** is proposed to better model the mRNA sequence. It largely improved the **SOTA up to 30%** on the downstream task.

**Predict the target site of MicroRNA** 9/2024 –

- mRNA sequence is too long to efficiently find the MicroRNA target site. I have developed an two stage miRNA and target mRNA learning method to overcome the interference from the mRNA redundant information

**Disease detection based on the Electrocardiograph (ECG) - Project Website** 10/2022 –

- I developed a machine learning model capable of evaluating the health of patients (age, disease) based on their electrocardiogram (ECG) data from the local hospital. The model can large **improve the accuracy of doctor's diagnosis**. The method has been **deployed** in the local hospital.

**Others** 08/2016 – 06/2019

- I Proposed an image super-resolution algorithm based on a micro-array camera (9x9); Proposed a real-time eye-tracking algorithm; Collaborated with North Tech Elderly Care to deploy a facial recognition system for the elderly.

**XtalPi Inc** 05/2024 – 10/2024
Computation R&D Internship Beijing
Analyzing the mRNA sequence and its representation. Developing the LLMs-based mRNA representation method and produce one academia paper.

**University of Hawaii at Manoa** 08/2019 – 12/2024
Research Assistant&Teaching Assistant Hawaii

**AMI Technology** 05/2019 – 07/2019
Machine learning Engineer Internship Beijing
Developed time series models for predicting Malaysian cell tower data. Responsible for applying classical time series algorithms, deep learning LSTM, and XGBoost algorithms to this dataset during the project.

**YouWin Education** 12/2015 – 08/2019
Part time High school mathematics tutoring.(Rank first in all mathematical teachers) Beijing

## Award

Giving Tree scholarship, University of Hawaii at Manoa 2024
Kola Innovation seeds, University of Hawaii at Manoa 2024
Outstanding Graduates of Beijing 2019
Second Prize of National Graduate Mathematical Modeling Competition (Captain) 2017
North China University of Technology academic star (1/150 in College of Science) ( 1/150 ) 2018
China National Scholarship 2018
Second Prize of China Undergraduate Mathematical Contest in Modeling (CUMCM) 2014
Second Prize of Mathematical Modeling League in the Three Northeast Provinces 2014

## Publication

**Papers**

1.Zhang, Honggen, et al. "Accurate prediction of age and sex using various machine learning models on 12-lead ECGs in children and young adults." (2024).
2.Zhang, Honggen, et al. "REAL: Response Embedding-based Alignment for LLMs." arXiv preprint arXiv:2409.17169 (2024).
3.Zhang, Honggen, et al. "mRNA2vec: mRNA Embedding with Language Model in the 5'UTR-CDS for mRNA Design." arXiv preprint arXiv:2408.09048, accepted by AAAI25, Philadelphia. (2024).
4. Zhang, Honggen, June Zhang, and Igor Molybog. "HaSa: Hardness and Structure-Aware Contrastive Knowledge Graph Embedding." Proceedings of the ACM on Web Conference 2024. 2024.
5. Zhang, Honggen, and June Zhang. "How News Evolves? Modeling News Text and Coverage using Graphs and Hawkes Process." arXiv preprint arXiv:2112.03008 (2021).
6. Zou, Jiancheng, and H. Zhang . New Key Point Detection Technology under Real Time Eye Tracking. Mechatronic Systems and Control.2019
7. Zou, Jiancheng, H. Zhang, and T. Weng. " Microarray Camera Image Super- Resolution with Neural Network and Fusion of V-System." International Conference on Computer Science and Education 2018

8. Zou, Jiancheng, and H. Zhang. ”Super-resolution reconstruction of color image based on microarray lens.” International Conference on Applied System Innovation IEEE, 2017:830-833.

## Skills

- Programming: Python > MATLAB > C++ > Java ≫ C#=MySQL
- Deep learning: Pytorch,TensorFlow, HPC Slurm
- Leadership: I organized two times (2023,2024) Chinese Booth section for the annual East-West Festival hold by east-west center.
- Courses:Doctoral Courses -Algorithm (Java); Algorithm Analysis (C++); Random Processes; Artificial Intelligence; Signal Processing; Convex Optimization; Mathematical Foundations of Reinforcement Learning; Computer Communication Networks. Master’s Courses-Digital Image Processing; Pattern Recognition; Mathematical Physics Equations. Undergraduate Courses-Fundamentals of Probability and Statistics; Mathematical Analysis; Advanced Algebra; Topology; Real Analysis and Functional Analysis.