
Haohong Zhang

Bioinformatics / Foundation models / AI for biology

Ph.D. student · Huazhong University of Science and Technology · Wuhan, China

haohongzh@gmail.com · ludenszhang.github.io · GitHub · Scholar

ORCID 0000-0001-6267-4244

Researcher working at the intersection of biological data and machine learning, with public code spanning transformer fine-tuning, discrete genome representations, microbial source tracking and scientific data visualization.

Technical skills

Machine learning. Python, PyTorch, PyTorch Lightning, Hugging Face Transformers and Datasets; masked language modeling, transfer learning and autoencoders.

Biological data. Biopython, FASTA processing, genome dataset preparation, microbiome abundance profiles and discrete sequence representations.

Analysis & tooling. NumPy, pandas, SciPy, scikit-learn, scikit-bio, plotnine; PCA/PCoA, command-line tools, Python packaging and TensorBoard.

Selected engineering projects

Nucleotide Transformer fine-tuning [Repository](#)

GTDB genome preprocessing and masked-language-model fine-tuning with Hugging Face Transformers and Datasets; configurable learning-rate schedules, tokenization and evaluation.

EXPERT-lightning [Repository](#)

Transfer learning for context-aware microbial source tracking, with a Python package, command-line interface and biome ontology.

Genome representation learning [Repository](#)

PyTorch discrete variational autoencoder training, reconstruction and regularization objectives, validation, early stopping and TensorBoard logging.

BriskyPCoA [Repository](#)

Command-line PCA and PCoA visualization of microbial abundance profiles using NumPy, pandas, SciPy, scikit-learn, scikit-bio and plotnine.

Education

2023 - Present Ph.D. in Bioinformatics · Direct-track

Huazhong University of Science and Technology · Advisor: Prof. Kang Ning

2019 - 2023 Bachelor of Science

Huazhong University of Science and Technology · Dengfeng Program, Class of 1901

Haohong Zhang

Research & engineering CV

Selected peer-reviewed research

MGM as a Large-Scale Pretrained Foundation Model for Microbiome Analyses in Diverse Contexts

H Zhang, Y Zhang, Z Kang, J Xiong, R Yang, K Ning · Advanced Science 13(24), e13333 (2026).

[10.1002/adv.202513333](https://doi.org/10.1002/adv.202513333)

Large-scale pretraining / Microbiome / Transfer learning

Deep learning enabled integration of tumor microenvironment microbial profiles and host gene expressions for interpretable survival subtyping in diverse types of cancers

H Zhang, X Xiong, M Cheng, L Ji, K Ning · mSystems 9(12), e01395-24 (2024).

[10.1128/msystems.01395-24](https://doi.org/10.1128/msystems.01395-24)

Multi-omics integration / Cancer survival / Interpretable deep learning

Preprints

MGM2 as a Unified Foundation Model for Microbiome World Exploration

H Zhang, Y Zhang, Y Qi, T Liu, R Yang, K Ning · bioRxiv (2026).

[10.64898/2026.07.20.739063](https://doi.org/10.64898/2026.07.20.739063)

Foundation models / Microbiome / Representation learning

MetaClaw: an auditable AI agent for end-to-end, multi-directional metagenomic and multi-omics analysis

H Zhang, Z Li, P N P Lagniton, Z Wang, L Zhao, W Li, P Duan, X Jiang, K Ning · bioRxiv (2026).

[10.64898/2026.07.21.739769](https://doi.org/10.64898/2026.07.21.739769)

AI agents / Metagenomics / Reproducible workflows

Conference presentation

A discrete token language of prokaryotic genomes learned by MicroVQVAE

Haohong Zhang, Kang Ning · Poster, Cold Spring Harbor Asia: AI and Biology.

April 20-24, 2026 · Suzhou, China.

Conference program · Poster 41

Honors

Outstanding Graduate, 2023 · Huazhong University of Science and Technology

Further research

Research software includes MGM2, MGM, MicroVQVAE, DeepMicroCancer and ASD-cancer. [Explore projects and full bibliography](#)