

HUILIN (HALEY) TAI

(646) 866-9171 · ht62@rice.edu · [linkedin.com/in/huilintai](https://www.linkedin.com/in/huilintai) · [haleyytai.github.io](https://github.com/haleyytai) · Houston, TX

My research focuses on interpreting, evaluating, and controlling multimodal foundation models, and their application to biomedical research.

EDUCATION

Rice University, Ph.D. in Computer Science Aug 2026 – Present
Advisor: Prof. Andrew Song (APEX Lab; joint with MD Anderson Cancer Center)

Columbia University, M.S. in Computer Science (Thesis Track), GPA 4.07/4.0 Sep 2024 – Dec 2025
Thesis: Cross-species antimicrobial-resistance prediction with genomic foundation models (Advisor: Prof. M. AlQuraishi)

McGill University, B.Sc. in Statistics and Computer Science, GPA 3.86/4.0 Sep 2020 – May 2024

RESEARCH EXPERIENCE

Rice University & MD Anderson Cancer Center, Ph.D. Researcher, APEX Lab Aug 2026 – Present

- **SP-Bench, multimodal foundation-model evaluation:** lead a unified benchmark of histology foundation models, six H&E-to-multiplex-IF generation models, and three spatial-proteomics foundation models on representation quality and downstream prediction; designed the modular inference adapters and evaluation protocols that let new models plug in without pipeline changes.
- **TigerCell, cross-modal prediction from pretrained vision transformers (with MIT/HMS):** predict cell-level gene expression from histology by pooling UNI-v2 ViT tokens over cell segmentations and fusing local kNN-transformer and global-attention context via per-cell, per-gene gating; ran 5-fold CV and image-encoder ablations against published baselines.
- **Theory of weak supervision:** derived information-theoretic lower bounds (Le Cam, Gilbert–Varshamov) showing an exponential separation between prediction and instance-level explanation recovery under bag-level labels; tested the predictions with pre-registered experiments on FEVER using five NLI readers.

Columbia University, ML Research Assistant, AlQuraishi Lab Aug 2024 – Dec 2025

- **Representation evaluation for a genomic foundation model (Evo):** built a cross-species antimicrobial-resistance prediction pipeline on intermediate transformer-block representations; layer-wise diagnostics (isotropy, effective rank, activation scale) identified layer 10 as the stable feature layer and exposed numerical instability in layers 11–12; MiniRocket aggregation over full-genome embeddings reached 95.5% cross-species accuracy.
- **Leakage-resistant OOD benchmark:** designed a leave-n-species-out evaluation comparing Evo-based models with the classic AMR predictor Kover, with automated hyperparameter tuning and MLflow tracking.
- **Memory-efficient inference:** block-wise compression, fp16, and HDF5 activation caching cut multi-GPU memory 10×, making full-genome (5M+ token) processing feasible.

Vision–Language Model Evaluation, Research Collaborator (Spec-LLaVA, T2VBench) 2024 – 2025

- **T2VBench (CVPR 2024 Workshop):** built the evaluation backbone combining VLM judges (LLaVA-1.5, InstructBLIP) with automatic metrics (CLIPScore, BLIPScore, VQAScore) over 1,600+ prompts and 5,000+ generated videos; ran correlation and annotator-consensus analyses against human preference.
- **Spec-LLaVA (ICML 2025 Workshop):** ran the large-scale experiments validating dynamic tree-based speculative decoding for LLaVA-1.5 (3.28× speedup); built the draft-vs-target evaluation and on-device scalability tests.

King Abdullah University of Science and Technology (KAUST), Research Associate May 2024 – Sep 2024

- Designed a generative video-editing architecture combining existing editing frameworks with a β -VAE for disentangled representations; implemented custom losses and training protocols in PyTorch.

Mila – Quebec AI Institute, Research Assistant, Ding Lab Dec 2022 – May 2024

- **CellSexID (first author, Cell Reports Methods 2025):** led an ML framework for tracking cell origin from single-cell transcriptomes; an XGBoost / RF / SVM / LR ensemble with Bayesian hyperparameter optimization reached 96% accuracy; released as a Python package and CLI.
- Designed a VAE with stochastic variational inference for spatial gene-expression modeling (custom ELBO, shifted-sigmoid transform); built multi-terabyte single-cell mining and clustering pipelines with cancer-center collaborators.

McGill University, Undergraduate Research Assistant (Profs. Hamed Hatami & Anmar Khadra) Nov 2022 – Apr 2024

- Designed a randomized group-testing protocol for Hamming-distance estimation, improving the communication upper bound from $O(\log n)$ to $O(\log \log n)$; studied replicability in agnostic learning; built MCMC models of nanoparticle–T-cell binding.

INDUSTRY EXPERIENCE

The Feinstein Institutes for Medical Research (Northwell Health), Data Scientist Intern May – Aug 2025; Feb – Jun 2026

- **Multimodal analysis pipeline:** led the end-to-end system linking H&E histology, spatial transcriptomics (13,000 spots), and immune gene signatures; used pathology foundation models to align tissue morphology with gene expression without task-specific training, and benchmarked spatial clustering methods (SpaGCN, Leiden, BANKSY, COVET) in a reproducible PyTorch pipeline on HPC (Docker, MLflow).
- **Grounded LLM reporting:** built an LLM layer that generates analysis reports cross-referenced against curated gene markers and biomarker rankings to reduce hallucination; paired it with a multi-algorithm biomarker-discovery framework with batch-effect correction and cross-animal consistency checks under small sample sizes.

Harbin Medical University, Statistician Intern

May 2022 – May 2023

- Survival analysis and transcriptomic modeling for hepatocellular carcinoma: built R-based nomograms for 1-, 3-, and 5-year survival prediction and ran GSEA (KEGG, Hallmark) across tumor risk groups; contributed to a Cellular & Molecular Biology Letters paper.

Yooden Information Technology, Data Science Intern (data center robotics)

Oct 2020 – Jan 2021

- Built a real-time streaming pipeline for multi-sensor robot data (SQL storage, ETL) with anomaly detection for power consumption and system performance; delivered Python analytics dashboards and automated load-balancing alerts.

PUBLICATIONS

Cross-Species Antimicrobial Resistance Prediction with Genomic Foundation Models

Huilin Tai (Advisor: Mohammed AlQuraishi)

arXiv preprint 2026 (M.S. Thesis, Columbia University)

CellSexID: Sex-Based Computational Tracking of Cellular Origins in Chimeric Models

Huilin Tai, Qian Li, Jingtao Wang, Jiahui Tan, Ryann Lang, Basil J. Petrof, Jun Ding [code]

Cell Reports Methods 2025

Spec-LLaVA: Accelerating Vision-Language Models with Dynamic Tree-Based Speculative Decoding

Mingxiao Huo, Jiayi Zhang, Hewei Wang, Jinfeng Xu, Zheyu Chen, **Huilin Tai**, Ian Yijun Chen

ICML 2025 Workshop

T2VBench: Benchmarking Temporal Dynamics for Text-to-Video Generation

Pengliang Ji, Chuyang Xiao, **Huilin Tai**, Mingxiao Huo

CVPR 2024 Workshop

Ependymal cells undergo an astrocyte-like gliosis in response to chronic and acute neuroinflammation

A. M. R. Groh, ..., **Huilin Tai**, J. Ding, A. Fournier, C. Larochelle, J. A. Stratton

Journal of Neurochemistry 2024

SPARC is a novel prognostic biomarker for ovarian cancer and associated with immune signatures and drug response

X. Guo, **Huilin Tai**, X. Li, P. Liu, J. Liu, S. Yu

Clinical and Experimental Obstetrics & Gynecology 2024

Targeting uridine-cytidine kinase 2 induced cell cycle arrest through dual mechanism and could improve the immune response of hepatocellular carcinoma

D. Wu, ..., **Huilin Tai**, ..., S. Tai

Cellular & Molecular Biology Letters 2022

SKILLS

ML / DL: PyTorch, Hugging Face Transformers, scikit-learn, XGBoost; foundation-model fine-tuning and representation extraction (ViTs, VLMs, LLMs, genomic LMs); benchmark and evaluation design (OOD splits, VLM-as-judge, statistical testing); sparse autoencoders and activation steering; multiple-instance learning; VAEs and diffusion models; Bayesian inference / MCMC

Systems: multi-GPU training and inference, mixed precision, Google Cloud Platform, HPC clusters, Docker, MLflow, HDF5

Languages: Python, SQL, R, Bash, C, Java, MATLAB

HONORS & TEACHING

1st Place, Intuit Explainable ML Hackathon

2026

Mackey-Glass Research Bursary, McGill Faculty of Medicine

2023

Hugh Brock Scholarship, McGill University

2020

Course Assistant, McGill University: Honours Probability, Statistical Learning, Discrete Math, Algebra

2022 – 2024