Kamal Maher

kamal.m.maher@gmail.com · kmaherx.github.io · github.com/kmaherx

Research interests: representation learning · interpretability · graph ML · biological foundation models

Education

2020 – 2025 | **MIT**, PhD in Computational Biology
Advisor: Xiao Wang

2014 – 2018 | **Cornell University**, BS in Neuroscience

Experience

2025 – 2026 | **SPAR**, Research Fellow
- Built an agentic framework that iteratively probes and explains learned representations in large language models, improving feature explanation accuracy by 7.7% over existing methods. [1]
- Demonstrated that a widely-used mechanistic interpretability method is incentivized to learn unfaithful explanations of neural network computation. [2]

2021 – 2025 | **MIT**, Graduate Researcher
- Developed a spectral graph framework linking tissue regions and cell–cell interactions via harmonic analysis of spatial transcriptomics. [3]
- Applied the framework to a 1.09M cell mouse brain atlas, revealing tissue boundaries absent from established references. [4]

2024 | **Genentech**, Aviv Regev Lab Intern
- Applied graph signal processing methods to cancer spatial transcriptomics data, identifying regions and cell–cell interactions more efficiently than existing GNN-based approaches.

Publications (*equal contribution)

[1] *Multi-shot AutoInterp: Agents Can Explain Complex Features By Refining Explanations.*
K. Maher*, S.E. Schrader*, K. Ayonrinde
ICLR 2026 Workshop on Unifying Concept Representation Learning. 2026
ICLR 2026 Workshop — From Human Cognition to AI Reasoning. 2026

[2] *Cross-Layer Transcoders are Incentivized to Learn Unfaithful Circuits.*
G. Lange, R. Goldstein, K. Dearstyne, **K. Maher**
LessWrong. 2026

[3] *Harmonic Representations of Regions and Interactions in Spatial Transcriptomics.*
K. Maher, X. Wang
bioRxiv. 2024

[4] *Spatial Atlas of the Mouse Central Nervous System at Molecular Resolution.*
H. Shi*, Y. He*, Y. Zhou*, J. Huang, **K. Maher**, B. Wang, Z. Tang, S. Luo, P. Tan, M. Wu, Z. Lin, J. Ren, Y. Thapa, X. Tang, K.Y. Chan, B.E. Deverman, H. Shen, A. Liu, J. Liu, X. Wang
Nature. 2023

Skills

ML / AI: representation learning · mechanistic interpretability · graph ML · agentic workflows
Tools: PyTorch · Git · LaTeX · Cloud (Runpod, Vast) · HPC (SLURM, UGE)
Biology: spatial transcriptomics · tissue biology · immunology · neuroscience