

ELLIOT TOWER

elliott@elliotttower.ai | elliotttower.ai | Boston, MA

EDUCATION

University of Massachusetts Amherst

Amherst, MA

M.S. in Computer Science — Concentration in Data Science

Sep. 2020 – May 2022

B.S. in Mathematics — Concentration in Computing, *Second major in Philosophy*

Sep. 2016 – May. 2020

EXPERIENCE

Visiting Researcher — University of Edinburgh, School of Informatics

Jun.. 2026 – Present

- Developed *sparse weight factorization*: decomposing transformers using a shared factor bank, exposing weight-level subspace communication channels using a shared interpretable vocabulary, a new level of granularity MI analysis.
- Developed per-factor circuit methods: Factorized EAP, Tucker/PCA decomposition on per-factor attribution tensor.

Machine Learning Engineer — Eluve Inc

Jul. 2024 – Mar. 2026

- Developed and deployed multilingual automatic speech recognition (ASR) + medical summarization pipeline for 13 specialties in Indonesian hospitals. Developed suite of evaluations for summary quality and hallucination detection.
- Developed tooling for longitudinal patient state and insights. Combined information extraction, retrieval, and medical knowledge graphs to produce alerts for drug-drug interaction, documentation gaps, and clinical decision support.

Machine Learning Engineer — Swarm Labs

May. 2023 – Jun. 2024

- Developed novel LLM evaluation benchmarks for adversarial robustness and red-teaming, AI safety and AI security.
- Benchmarked frontier and open-source LLMs, discovered novel failure modes, and created model scorecard reports.

Open-Source Project Manager — Farama Foundation

Mar. 2023 – Jun. 2024

- Project manager of PettingZoo, the API standard for multi-agent RL. Developed ML training integrations and tooling.

PROJECTS

Mechanistic Validity Framework [\[Zenodo\]](#) [\[Website\]](#)

Mar. 2026 – Present

- Developed validity framework for mechanistic claims, extending philosophy of mechanisms and measurement theory
- Unified validation methodology from philosophy of science, causal inference, neuroscience, pharmacology, genetics
- **Applied to mechanistic interpretability**: audited 13 published results (IOI, SAE, probing), identified systemic gaps.

Mechanistic Research (Research Program) [\[Website\]](#)

Mar. 2026 – Present

- Geometric methods for predicting when mechanisms transport across domains, populations, and instruments.
- **Bracket Norm**: Identified causally important brain regions against ground truth optogenetic silencing, using novel geometric method based on vector field lie algebra. All standard methods confounded by neuron count. [\[Zenodo\]](#).
- **Direction Instability**: novel metric for predicting drug mechanism transport across cell lines, tested on transcriptomic [\[Zenodo\]](#), genetic perturbation and cell painting morphology [\[Zenodo\]](#). Practitioner guide. [\[Zenodo\]](#)
- **Genetics Cis-pQTL** Mendelian randomization uninformative for enzyme inhibitors and receptor antagonists [\[Zenodo\]](#)
- **RNA/DNA**: 8/10 foundation models do not encode structure from nucleotide composition (MI evaluation) [\[Zenodo\]](#)
- **Quantum NV-diamond Sensors**: Geometric invariance of a sensor metric predicts its blind spots [\[Zenodo\]](#)

PUBLICATIONS

Das R, Godbole A, Naik A, Tower E, Zaheer M, Hajjishirzi H, Jia R, McCallum A. **Knowledge Base Question Answering by Case-based Reasoning over Subgraphs**. ICML, 2022. [\[Proceedings\]](#)