ELLIOT TOWER

elliot@elliottower.ai | elliottower.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.

RESEARCH

Mechanistic Validity Framework [Zenodo] — Mar. 2026 – Present

- Developed domain-general validity framework for the triangulation and evidence formalization of mechanistic claims.
- Unified validation methodology from philosophy of science, causal inference, neuroscience, pharmacology, genetics
- **Applied to mechanistic interpretability:** audited 16 published results (IOI, SAE, probing), identified systemic gaps.

Mechanistic Research (Research Program) [Website] — Mar. 2026 – Present

- Geometric methods for predicting mechanism transport across domains. Mechanistic interpretability for science.
- Methodology: pre-registration, permutation null controls, confound-first experimental design, Zenodo deposit results.
- **Single-cell:** foundation model embeddings ranked at chance by 5/5 existing metrics, derived analytic failure bound, validated replacement metric, transfer to downstream tasks (ρ = 0.67, 25 tissues), open-source library [Zenodo]
- **RNA/DNA**: 8/10 foundation models have no structural awareness after controlling for nucleotide composition confound; graded three-rung evaluation across 52 Rfam families. RiNALMo (650M) best, Evo (7B) weakest [Zenodo]
- **Pharmacology:** developed novel metric for predicting drug mechanism transport across cell lines, tested on transcriptomic [Zenodo], genetic perturbation and cell painting morphology [Zenodo]. Practitioner guide. [Zenodo].
- **In silico genetics:** higher-order knockout epistasis is unpredictable from molecular wiring; exhaustive 2^n sweeps across 28 Boolean networks, strongest local interactions produce weakest global effects, ODE-validated. [Zenodo]

PUBLICATIONS

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