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

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 Intern — Information Extraction and Synthesis Laboratory (IESL) — Jun. 2021 – Aug. 2021

- Collaborated in developing novel architecture using nonparametric case-based reasoning and graph neural networks.
- Publication (co-author): *Knowledge Base Question Answering by Case-based Reasoning over Subgraphs* **(ICML 2022).**

Industry Mentorship — UMass Amherst & Facebook AI Research (FAIR) — Feb. 2021 – Jun. 2021

- Implemented, trained, and benchmarked graph transformers for molecular chemistry tasks (Open Catalyst Project).

RESEARCH PROJECTS

Mechanistic Validity Framework — Mar. 2026 – Present

- Developed novel framework extending mechanistic interpretability, with a full 5-layer taxonomy for validating claims.
- Adapted validation methodology from neuroscience, philosophy of science, pharmacology and measurement theory.
- Analyzed 13 published results (IOI, SAEs, probing) to find systematic gaps in causal evidence and construct validity.

Taxonomy of LLM Reasoning Failures — Mar. 2026 – Present

- Developed suite of reasoning and self-evaluation tasks, exposing four novel failure modes across frontier models.
- Applied causal inference methods and ran MI ablations/analysis to isolate four structurally distinct mechanisms.
- Developed novel solutions for each failure mode, benchmarking scaffolds for mid-generation detection & mitigation.

Clinical Hallucination Detection & Security Vulnerability Detection — Mar. 2026 – Present

- Benchmarked 20+ detection strategies: prompting, retrieval-based, multi-step, self-correction, and model consensus.
- Ran 2x2 causal factorial design with OLS, causal forest, and bootstrap resampling, across 6 models and 15 datasets.
- Uncovered binary verification framing as causal mechanism for false negatives for both security and medical tasks.
- Developed solution using task restructuring, and improved performance with domain-specific multi-step pipelines.