Aravinth Ravi arav0006@ksu.edu | https://hippyhippohops.github.io/ | github.com/hippyhippohops

Education

Kansas State University

- **PhD** in Applied Mathematics. Aug 2022 – May 2027
- **Research Topic:** Developing deep learning algorithms for inverse problems. **Advisor:** Dinh-Liem Nguyen.
- *Friends of Mathematics Award in Research* – Awarded to top graduating student for research 2026–2027

Nanyang Technological University

- **B.Sc. (Honours with Distinction)** in Mathematical Sciences. Aug 2018 – June 2022
- *President Research Scholar* – Awarded to top-performing students for the undergraduate research 2019–2021

AI Safety Projects

Core Oscillation in Local Factor Graph Debate (replication) Current

- *Local Factor Graph Debate* (LessWrong, 2026). On 600 random graphs with one cycle, found the self-avoiding-walk upper bound on core oscillation is a median 2.3–2.5× above the exact value. Computed exact oscillation in two independent ways: brute-force enumeration on small pairwise factor graphs, and an exact belief-propagation recursion with interval-valued messages on trees. The two agree to within 10^{-13} on 500 random trees. Link to notebook.

Belief-State Geometry in Transformers, BlueDot Technical AI Safety Project Sprint Current

- Worked through the theory of Shai et al. (NeurIPS 2024) on a hidden Markov model end to end: derived the Bayesian belief-update rule and verified it against the paper's Figure 3. Reproduced the Z1R (link to notebook) and Mess3 (link to notebook) experiment, training a small transformer using TransformerLens on HMM-generated tokens and testing by linear regression whether its residual stream encodes the predicted belief-state geometry.

OpenAI's AI Agent Security - Multi-Step Tool Attacks Kaggle Competition Summer 2026

- Implemented various deterministic red-teaming multi-step attack search algorithms (including a probe-based template race algorithm and Go-Explore and Rainbow-teaming-style attack search algorithms) that drive tool-using LLM agents (GPT-OSS 20B, Gemma 4 26B-A4B) across safety guardrails into unsafe tool calls (data exfiltration, destructive writing).

Academic Experiences

Graduate Research Assistant, Kansas State University Fall 2022 – Current

- Simulated Coulomb explosions by integrating coupled equations of motion and trained MLP and VAE models to reconstruct 3D molecular structures from fragment momenta; the VAE reached $R^2 \approx 0.99$ on held-out samples. (*J. Chem. Phys.* **164**, 244304, 2026.)
- Designed a Fourier-domain model loss function for the multifrequency Helmholtz inverse source problems and integrated this to construct an unsupervised model-informed deep learning algorithm; reduced 3D convolution evaluation from $\mathcal{O}(N^6)$ to $\mathcal{O}(N^3 \log N)$ per frequency using inverse FFTs, recovering source coefficients from boundary measurements with 10% noise. (arXiv:2609.33062.)
- Derived explicit Fourier coefficients and proved decay estimates for periodized Helmholtz Green kernels; developed a spectral physics-informed deep learning algorithm jointly estimating scattering contrast and contrast source through coupled Lippmann–Schwinger and imaging equations, reconstructing 2D and 3D scatterers from noisy, phaseless boundary measurements. (arXiv:2609.33070.)

Undergraduate Research Assistant, Nanyang Technological University 2019–2022

- Built a pipeline to test Lê's conjectured volume lower bound on the virtual homological spectral radius of pseudo-Anosov maps. Enumerated 266,256 candidate covers of the once-punctured torus, computed the homology action of the 265 valid lifts via Fox calculus, and found the bound held in every case (2.618 against a required 1.240). (Thesis).

Technical Skills & Professional Activities

- *Service:* Reviewer for NeurIPS 2026 Interpretability for Discovery workshop
- *Competitive Programming:* Codeforces (Specialist Rank)
- *Languages and tools:* Python, MATLAB, LaTeX, Claude Code, Git, Linux (SLURM).
- *Libraries & Frameworks:* PyTorch, scikit-learn, pandas, NumPy, Hugging Face, TransformerLens.
- *Coursework:* An Introduction to Statistical Learning, Mathematics of Deep Learning, AI and Mathematics, Deep Learning, Pattern Recognition & Machine Learning, The Mathematics of Data and Networks (probabilistic graph theory) .
- *Certifications:* BlueDot AGI Strategy course, BlueDot Technical AI Safety course
- *Reading Groups:* BlueDot AI Safety Evals Paper Reading Club, Universal Algorithmic Intelligence reading group