Theodore Ehrenborg theodore.ehrenborg@gmail.com ehrenborg.dev

Work

Dec 2025 – Present Research Engineer, Principles of Intelligence (PIBBSS)
Investigating a novel sparse autoencoder architecture

May 2025 – Nov 2025 Researcher, Beneficial AI Foundation
Used formal verification to check AI-written software

Feb 2025 – Mar 2025 Technical AI Safety Fellow, Pivotal Research. Mentor: Logan Riggs Smith

Jan 2024 – Jan 2025 Senior Machine Learning Engineer, Myrtle.ai, Cambridge, UK
For our CAIMAN-ASR product, I
- stayed current with the speech recognition literature
- ran experiments to reduce word error rate. See `ehrenborg.dev/caiman`
- debugged hardware problems in our on-premise GPU servers
- interviewed applicants and interacted with potential customers

Sept 2022 – Dec 2023 Junior Machine Learning Engineer, Myrtle.ai

Summer 2022 Machine Learning Intern, Myrtle.ai
Designed voice activity detector, created dataset, and trained LSTM RNNs.

Research

March 2025 Discovered symlink exploit in Vast.ai. Awarded bug bounty. See `ehrenborg.dev/exploit`

Spring 2025 Mechanistic interpretability project at Pivotal Research. See `ehrenborg.dev/absorption`

April 2025 Cowrote grant proposal, which Coefficient Giving funded. See `ehrenborg.dev/sleeper`

Sep 2025 Second author on "A benchmark for vericoding". See `arXiv:2509.22908`

GitHub `github.com/TheodoreEhrenborg`

Author of `sae.ehrenborg.dev`
Trains a sparse autoencoder on a 33 million parameter language model (LM)

Author of `rl.ehrenborg.dev`
Uses reinforcement learning to make a LM generate stories with alliteration

GitHub while at Myrtle.ai `github.com/TheoEhrenborg`

Contributor to `github.com/MyrtleSoftware/caiman-asr`
Author of 412 of the 926 merged pull requests, which include:
- Increasing long-audio validation speed by 8x
- Random State Passing, which decreased long-audio word-error-rate by ~40% relative

Reviewed 382 of the other 514 merged pull requests. Released v1.9.0, v1.10.1, and v1.11.0

Education

2019 – 2022 University of Cambridge, Cambridge, England
BA Mathematics, Clare College. Awarded Cambridge Trust Scholarship, £10000 per year.
Year 1: Not classed due to pandemic. Year 2: Class I. Year 3: Class II.i.

2015 – 2019 Henry Clay High School, Lexington, KY, USA
SAT score: 1600 out of 1600. ACT score: 36 out of 36. 13 AP courses: all scores 5 out of 5.

Computer Experience

Fluent in: Python, PyTorch, Docker, Git, GitHub Actions, Linux, WordPress, & LaTeX
Have also used for work: Rust, Nix, Elm, AWS, Compute Engine, Vast.ai, Hydra CI, Dafny, Verus, & Rocq
Have also used: Lean 4, Julia, Java, Keras, SageMath, & Octave

Extracurriculars

Winter 2019 – Spring 2020 Imperial College Mathematics Competition, tied for 43rd place nationally

May 2018 Intel International Science and Engineering Fair,
"Pythagorean Quintuples and Quaternions" won a 3rd award in math