

HENRY FREED

(206)-617-2572 ◊ henry@hbfreed.com ◊ hbfreed.com

PROJECTS

nanoMoE

Variable Sized Mixture-of-Experts

- Designed and managed end-to-end research program studying compute allocation in MoE language models, from hypothesis development through evaluation methodology to published findings.
- Built evaluation framework measuring token routing patterns across domains; discovered routing correlates with contextual ambiguity, not computational difficulty—code tokens route 14σ below cross-domain mean.
- Validated reproducibility across multiple random seeds, identifying that 30% of routing decisions show stable learned structure—establishing measurable signal from noise.
- Extended nanoGPT with MegaBlocks sparse kernels to support variable-sized expert configurations at 125M–250M parameter scale.

Variable FlexOlmo

Variable-Width Expert Pruning for FlexOlmo

- Pruned Ai2's FlexOlmo math expert to three widths (74%, 50%, 19%) using importance-based neuron scoring and KLD distillation on 228M tokens. The 74%-width expert matches baseline; the 19%-width expert scores $3.6\times$ above the no-expert model.
- Designed evaluation comparing domain-specific vs. general importance analysis for pruning; found 58% neuron disagreement between methods, with domain-calibrated pruning producing measurably better training dynamics.
- Released logprobs dataset and trained model weights on HuggingFace.

Open Concept Steering

Mechanistic Interpretability

- Developed and open-sourced a method for extracting interpretable concepts from language models using sparse autoencoders (SAEs), training a 65k-dimensional SAE on 600M activations from OLMo-2-7B.
- Released full reproducibility package: interactive demo, blog post, dataset, training code, and model weights.

EMPLOYMENT

Cirtec Medical, Volant Partners

June 2024 - September 2025

AI/ML Engineer - Generative AI (Contract)

Remote

- Owned full project lifecycle as sole AI technical resource: gathered requirements from marketing leadership, scoped solutions, built and deployed systems, and defined quality metrics—all on 1–2 week delivery cycles.
- **Cirtec Chat** – Designed RAG-based knowledge system end-to-end. Implemented retrieval pipeline, built evaluation framework (95% recall), and created usage analytics dashboard to track quality and inform iteration.
- **AI Content Pipeline** – Architected cross-platform automation system, aligning marketing, brand, and technical requirements into a consistent content delivery pipeline.
- Translated qualitative stakeholder feedback into technical requirements and measurable success criteria across all projects.

VanEck Securities*Emerging Markets Fixed Income Intern*

June - September 2022

New York, NY

- Built Python tooling to automate sales presentation generation, reducing task completion time from 30 minutes to 3 minutes.

Portland Pickles Baseball Team*Assistant Director of Operations*

July - August 2021

Portland, OR

- Implemented Kanban system for pre-game operations; managed interns and oversaw game-day logistics.
- Collected and tracked player performance metrics (exit velocity) using Trackman; managed live-stream production.

Naka and Adana Restaurants*Demi Sous Chef*

February 2016 - November 2017

Seattle, WA

- Led kitchen operations under chef Shota Nakajima; managed transition from Naka to Adana including developing and deploying an entirely new dessert program on a compressed timeline.

Meld*Product Management Intern*

June 2015 - August 2015

Seattle, WA

- Developed systematic process for recipe testing and integration: identified candidates, ran tests, gathered cross-functional feedback, and iterated to meet product quality standards.

TECHNICAL SKILLS

Languages & ToolsPython (PyTorch, HuggingFace Transformers, pandas, NumPy, OpenAI SDK), L^AT_EX, Git**LLM & ML**

Training, Distillation, Inference, Evaluation Pipelines, MoE Architectures, Sparse Autoencoders, RAG Systems

Evaluation

LM Evaluation Harness, Benchmark Design, Retrieval Metrics, Usage Analytics

EDUCATION

Portland State University*B.S. Applied Mathematics*

Graduated August 2023

*Portland, OR***INTERESTS**

- Science fiction (Liu Cixin's *Remembrance of Earth's Past*, Stephenson's *Snow Crash*), baseball, home-brewing
- Biking, weightlifting, cooking, the Grateful Dead, pawpaw fruit enthusiast