Hugh Flournoy Van Deventer V

1-917-685-3090 hughvandeventer@g.harvard.edu linkedin.com/in/hughvandeventer github.com/hughvd

EDUCATION

Harvard University **Expected May 2027**
S.M. Data Science. GPA 3.8 *Cambridge, MA*
Relevant Coursework: AI Safety & Alignment, MLOps, Intro to Linear Models, Intro to Data Science.
Activities: AI Safety Student Team.

University of Michigan **May 2025**
B.S. Mathematics and Physics, Minor in Computer Science. GPA 3.8 *Ann Arbor, MI*
Relevant Coursework: Machine Learning, Continuous Optimization, Linear Programming, Numerical Linear Algebra, Science of LLMs, Complex Systems Modeling, Computational Physics, Probability Theory.
Activities: Physics Department Peer Advisor, Athletics Department Tutor, Wolverine Tutor, Overwatch Team Captain.

Bronx High School of Science **June 2021**
GPA: 4.0 *New York, NY*

RESEARCH EXPERIENCE

Cambridge Boston Alignment Initiative **Feb. 2026 – Present**
Research Fellow *Remote*
- Selected from ∼300 applicants for the CBAI Spring 2026 Research Fellowship to work with Dr. Laura Ruis, with an initial focus on **generalization and out-of-context reasoning (OOCR)** in language models.

Harvard University **Sep. 2025 – Present**
Research Assistant *Cambridge, MA*
- Developing extensions of **masked diffusion models** utilizing ideas from **recursive reasoning** frameworks.

University of Michigan Center for Academic Innovation **Oct. 2023 – Aug. 2025**
Data Science Fellow *Ann Arbor, MI*
- Developed novel two-stage **RAG methodology** addressing semantic gap in retrieval by generating ideal descriptions as intermediate query representations, improving course recommendation relevance over direct embedding similarity.
- Built and deployed **LLM-powered course recommender** (**FastAPI**, **AWS**) serving **10K+ courses** to university community, with bias analysis and network visualization validating embedding space semantic relationships across academic domains.
- Led research culminating in **first-author publication** and collaboration with Michigan Online for professional certification recommendations.

Michigan Tech Research Institute **May. 2024 – Aug. 2024**
Machine Learning Research Intern *Ann Arbor, MI*
- Led literature review on ML for **super resolution and image registration** for a Ford automotive camera project.
- Designed and trained a CNN with a custom loss function to predict warping parameters for 128x128 image chips, reducing LBFGS optimizer iterations by **30%** and accelerating image registration processing times.
- Implemented framework enabling custom gradients for functions incompatible with MATLAB autodifferentiation.

INDUSTRY EXPERIENCE

Honeywell **Jun. 2025 – Aug. 2025**
AI/ML Engineering Intern *Atlanta, GA*
- Developed **computer vision** and **OCR**-based floor plan digitization system for automated corporate building bidding pipeline, enabling downstream sensor placement optimization and cost estimation.
- Led R&D evaluation of open-source, academic, and foundation models for floor plan semantic segmentation.
- Built an **agentic vibration analysis system** for industrial equipment that autonomously diagnoses mechanical faults by combining signal processing with LLM-powered expert reasoning.

Neurabuild **Jul. 2023 – Aug. 2023**
Machine Learning Intern *Capetown, South Africa*
- Developed ML solutions to automate sky visibility for portable astronomical sites, including a W-net for semantic segmentation of clear vs. cloudy skies and a CNN classifier achieving **95%** accuracy in night sky condition detection.
- Improved existing neuromorphic satellite detection and tracking model performance by **10%** using edge detection and KerasTuner for hyperparameter optimization.

PROJECTS

Are Personas All You Need? Stress-Testing Persona Vectors as Alignment Tools | *AI Safety, Interpretability, Evaluation*

- Evaluating whether **persona vectors** are reliable tools for **monitoring alignment drift** during fine-tuning. Found persona similarity correlates with **cross-trait effects** at a coarse level, but robustness failures are common.

HistoriCam: Photo-to-History Landmark Identification | *MLOps, Computer Vision, RAG*

- Built a web app that identifies a building from a photo and retrieves historical context; implemented the **building identification pipeline** using vision embeddings and **vector search** with top-k majority voting above a cosine-similarity threshold. Deployed the vision + LLM backend on **GCP Kubernetes** with autoscaling.

Prompting Our Way to Safety: Comparing System Prompt Strategies | *Huggingface, AI Safety, Evaluation*

- Conducted experiment for Harvard CS 2881r comparing system prompt strategies on Deepseek-R1 models; found principled prompts can achieve safety comparable to fine-tuned models while reducing over-refusal by 2-10×. Published on LessWrong.

Unembedding Steering in Large Language Models | *PyTorch, Transformers, Representation Engineering*

- Developed novel steering method using averaged token unembedding vectors, comparing against linear probing and CAA on Gemma-2-2b with contrastive evaluation framework. Demonstrated competitive performance with learned methods while requiring no training, revealing alignment between token representations and internal model activations.

optiMaizer: Optimization Algorithm Benchmarking Suite | *Python, Numerical Optimization, Performance Analysis*

- Implemented and benchmarked 10 optimization algorithms across diverse test functions. Conducted systematic L-BFGS memory analysis revealing intermediate memory sizes often outperform large configurations on complex landscapes.

SHLIME: Adversarial Robustness for Explainable AI | *Python, Adversarial ML, XAI*

- Replicated adversarial attacks against LIME and SHAP, developing novel combined defense method improving robustness while maintaining explanation quality.

TECHNICAL SKILLS

Languages: Python, C++, SQL, TypeScript, MATLAB, LaTeX
Machine Learning: PyTorch, TensorFlow, Transformers, Hugging Face, Scikit-learn, XGBoost
ML Engineering & Production Systems: OpenAI API, LangChain, LLM Agents, Vector Databases / RAG Systems, Model Distillation & Quantization, OpenCV, Docker, Kubernetes, MLflow, Weights & Biases, Git, FastAPI
Cloud & Infrastructure: AWS, Google Cloud, Azure ML, CI/CD, Data Pipelines & ETL
Data & Visualization: NumPy, Pandas, SciPy, Matplotlib, Plotly, Jupyter