

Alexander Loftus

I am an AI researcher and strong communicator with deep technical expertise. I strive to help humans understand, evaluate, and steer AI more meaningfully, powerfully, and safely. My work has spanned representation learning, scalable data processing pipelines, volumetric segmentation, and network statistics.

Some highlights:

1st place ranking, \$100,000 Machine Learning competition: [Work featured on cover of Scientific American](#). Competed against 1249 teams. Vesuvius scroll ink detection.

Textbook author: Publishing contract with Cambridge University Press.

Teaching and Leadership: Assistant director managing 10 team members; led a team of three to develop an object detection augmentation algorithm; led a team of five to contribute to a brain network estimation pipeline.

Publications in top conferences: Best poster award at NeurIPS 2023 LatinX workshop, first author work in ICLR 2024.

Open-source contributions: Primary contributor to microsoft network statistics package graspologic.

TEXTBOOK

Hands-on Network Machine Learning with Python: *Eric Bridgeford, Alexander R. Loftus, Joshua Vogelstein.*

Cambridge University Press. Publication August 2025.

Spectral representation theory on networks. 530 pages, 147 figures.

EXPERIENCE

Data Scientist

Creyon Bio

San Diego, CA

2022-2024

ESP Embeddings: Developed a novel contrastive learning method for 3D electrostatic potential data, classifying toxicity for early drug screening.

Neuron Toxicity Detection: Built deconvolution and segmentation pipeline to detect toxicity in neurons. Resulted in a new drug discovery pipeline.

Machine Learning Research Engineer

Blue Halo

Rockville, MD

2021-2022

Conditional Image Generation with Generative Adversarial Networks: Spearheaded synthetic data generation with diffusion models, replacing less performant GAN-based workflows.

Detecting Objects with Enhanced Yolo and Knowledge Graphs: Designed method for conditional generation and object detection on large video datasets using enhanced YOLO and knowledge graphs.

Geometric Multi-Resolution Analysis: Infrastructure for a hierarchical clustering method.

Artificial Intelligence Research Engineer

Johns Hopkins University — Dr. Joshua Vogelstein

Baltimore, MD

2018-2021

Graspologic: Co-authored an open-source graph statistics library later adopted by Microsoft Research.

ndmg: Optimized a diffusion MRI pipeline with Kubernetes and AWS orchestration, cutting 1000 lines of legacy code and halving runtime.

Assistant Director

iD Tech Camps — University of Washington

Seattle, WA

2014-2018 summers

Leader and Manager: Directed a high-throughput STEM education camp. Managed 10+ instructors weekly and developed curriculum in C++, Python, and robotics.

Curriculum Designer: Designed and deployed teaching tools and live demos across 50+ locations.

EDUCATION

Northeastern University

PhD Computer Science

Advisor: [Dr. David Bau](#)

Boston, MA

2024-

Interpretability, evaluations, and training dynamics in Code LLMs.

Johns Hopkins University

MSE Biomedical Engineering: Machine Learning & Data Science Focus

Advisor: Dr. Joshua Vogelstein

Thesis: Hands-On Network Machine Learning

Baltimore, MD

2020-2022

dean's list, highest honors, GPA 4.0/4.0.

Western Washington University

BS Behavioral Neuroscience — Minors: Chemistry, Philosophy

Founder & President, Computational Neuroscience Club

Vice President, Neuroscience Club

Bellingham, WA

2014-2018

Built computational neuroscience club from scratch, taught weekly seminars.

PUBLICATIONS

* indicates equal contribution.

🏆 indicates best poster.

NNsight and NDIF: Democratizing Access to Open-Weight Foundation Model Internals: *A.R. Loftus**, *J. Fiotto-Kaufman**, et al. ICLR 2025.

Open-source suite for probing & manipulating LLM weights without engineering overhead. Ray GCS Service backend with AWS object storage and VLLM for inference speed.

🏆 A Saliency-based Clustering Framework for Identifying Aberrant Predictions : *A. Tersol Montserrat, A.R. Loftus, Y. Daihes*. Paper, **NeurIPS LatinX AI Workshop**, 2023.

Detects spurious feature reliance via saliency embeddings.

A low-resource reliable pipeline to democratize multi-modal connectome estimation and analysis: *J. Chung, R. Lawrence, A.R. Loftus, et al*. Paper, in review at Nature Methods, 2024

Transforms diffusion MRI scans into graphs; open-sourced ([code](#))

TALKS

NNsight and NDIF: *A.R. Loftus*, FAR AI conference 2025

Talk+Demo on scaling interpretability techniques to large foundation models. Technical policy conference in Washington, DC.

State of the Art in Knowledge Editing: *A.R. Loftus*, 2023

Current techniques in knowledge localization and editing in LLMs and diffusion models.

1st Place Solution - Vesuvius Ink Competition: *R. Chesler, A.R. Loftus, A. Tersol Montserrat, T. Kyi*, 2023

Presenting on our winning solution to a \$100,000 Kaggle competition, part of the \$1,000,000 Vesuvius competition.

ICML Conference Highlights: *A.R. Loftus*, 2023

Machine learning techniques in drug discovery and medicine at ICML 2023.

Working with LLMs: *A.R. Loftus*, 2023.

Introduction to LLM engineering. Talk given to 100 people at the AI/ML San Diego meetup.

Linear Algebra, from Dot Products to Neural Networks: *A.R. Loftus*, 2023.

Created a YouTube tutorial series on the fundamentals of linear algebra for machine learning.

Effects of an unc-43 (CaMKII) Gene Deletion on Short-Term Memory for Associative Conditioning in C. elegans: *A.R. Loftus*, Psychfest 2017.

Mechanistic understanding of worm neural circuitry.

FELLOWSHIPS & AWARDS

First Place Winner

Kaggle Vesuvius Competition, \$100,000.

2023

Khoury Distinguished Fellowship

Northeastern University PhD fellowship.

2024

Best Poster Award

NeurIPS 2023 LatinX AI Workshop.

2023

MIT EECS GAAP

MIT mentorship program.

2023

| | |
|---|------|
| Harvard AI Safety Technical Fellowship Harvard fellowship. | 2025 |
| AWS Research Grant \$10,000 grant for computational research on cloud services. | 2019 |

SERVICE & ADVISORY

| | |
|---|-------------------------|
| Consultant Advisory role for cybersecurity-based interpretability startup. | Krnel.ai Spring 2025 |
| Organizer Organizing the 200-person New England Mechanistic Interpretability conference. Obtained \$17,000 grant. | NEMI Summer 2025 |
| Mentor Harvard/MIT joint program designed to support early-stage researchers. | CBAI Summer 2025 |

TEACHING

| | |
|--|--|
| Head Teaching Assistant Foundations of Computational Biology and Bioinformatics, <i>EN.BME.410/634</i> | Johns Hopkins University Spring 2021 |
| Teaching Assistant NeuroData Design II , <i>EN.BME.438/638</i> | Johns Hopkins University Spring 2020 |
| Teaching Assistant NeuroData Design I , <i>EN.BME.437/637</i> | Johns Hopkins University Fall 2019 |
| Teaching Assistant Introduction to Behavioral Neuroscience, <i>PSY.220</i> | Western Washington University Winter 2017 |
| Curriculum Designer Built curriculum used across 50 locations in the United States by tens of thousands of students. | iD Tech Camps Spring 2017 |
| Instructor Taught programming and game design to high school students. | iD Tech Camps 2014-2018 summers |

SKILLS SUMMARY

Languages: Python, Bash, R, Rust, SQL

Tools & Frameworks: pytorch, pytorch-lightning, tensorflow, jax, numpy, scipy, pandas, polars, sklearn, seaborn, matplotlib, docker, AWS, google cloud (GCP), photoshop, SQL, weights & biases, mlflow, kubernetes, linux

Areas of Expertise: Linear algebra, probability & statistics, deep learning, information theory, transformers, diffusion models, convolutional autoencoders, GPUs and CUDA, public speaking, leadership & management, teaching, natural language processing, computer vision

Soft Skills: Scientific writing, public speaking (100+ audience LLM talks), leadership, rapid debugging, knowledge distillation

FUN

Gaming: Starcraft 2 grandmaster in high school, competed and won in Seattle-area tournaments.

Music: Fingerstyle guitarist. Played at open mic nights.

Dancing: Partner dance instructor and competition winner. Fusion, West Coast Swing, Zouk, Salsa, Bachatta.