

Aidan Erickson

aidanerickson1@gmail.com ♦ github.com/aidan-sonoferick ♦ linkedin.com/in/aidan-sonoferick/

EDUCATION

Carnegie Mellon University

Master of Science: Artificial Intelligence

GPA: 3.78/4.0

Coursework: Generative AI | Deep Learning | Speech Recognition | Natural language processing | Systems and Toolchains for AI

Michigan State University

Bachelor of Science: Computer Science, Minor in Mathematics

Pittsburgh, PA

Graduated 12/2024

East Lansing, MI

Graduated 5/2023

PROFESSIONAL EXPERIENCE

Wynd Labs / Grass - Machine Learning Research Engineer

New York, NY | 04/2025 - Present

- Designed highly optimized and scaled machine learning annotation pipelines, processing billions of textual, audio, and video data for leading frontier labs. Created high-throughput ML pipelines in only a few days for quick customer turnaround.
- Trained Cliptagger, a fine tune of Gemma3-12b, beating Claude 4 Sonnet at video keyframe annotation at 1/20th the price.
- Designed webpage metrics and ranking system for AI usefulness. Deployed across 60k high-signal websites in 59 countries.
- Created multimodal retrieval systems from internet-scale webcrawls (billions of embeddings) with our Grass proxy network.

Besty - Founding Engineer

New York, NY | 02/2025 - 04/2025

- Created demand forecasting and dynamic pricing models to optimize stay prices on over 6M reservations on 50k listings.
- Designed, created, and stress tested system integrations for customers with limited documentation and 3-day deadlines.
- Developed test environments to simulate high-load onboarding before making the platform publicly accessible.

Visa, Automated Payment Fraud Disruption - Machine Learning Engineering Intern

Bellevue, WA | 05/2024 - 08/2024

- Developed a convolutional graph neural network to detect identity fraud at scale (~2.6B Entries), tightening fraud analysis.
- Conducted experiments and research on different methodologies, including transformer and convolutional recurrent models, among others, achieving an identity fraud prediction accuracy of over 89%. Used Neo4J and PyTorch Geometric on CUDA.

Amazon, Project Kuiper - Software Development Engineering Intern

Seattle, WA | 05/2022 - 08/2022

- Implemented asset configuration features in GoLang, affecting 3,200 low orbit broadband satellites and ground equipment.

PUBLICATIONS

Project Alexandria 2.0: Mapping 22 Million Scientific Papers through LLM-Distilled Structured Knowledge Sub. Neurips 2026

Christoph Schuhmann, Gollam Rabby, Robert Kaczmarczyk, **Aidan C. Erickson**, Van Khue Nguyen, Harsh Raj, Andrii Prolorenzo, Andrej Radonjic, Huu Nguyen, Victor May, Ajinkya Bawase, Sören Auer, Jenia Jitsev

LAION-Tunes: An Open 1.4M-Track Dataset and Perceptual Benchmark for AI-Generated Music

Sub. Neurips 2026

Robert Kaczmarczyk, Tawsif Ahmed, Felix Friedrich, **Aidan C. Erickson**, Orian Sharoni, Dorien Herremans, Christoph Schuhmann

Toward a Low-Cost Perception System in Autonomous Vehicles: A Spectrum Learning Approach

IEEE ITSC 2025

Mohammed Alsakabi, **Aidan C. Erickson**, John M. Dolan, Ozan K. Tonguz

VOLUNTEERING

Large-scale Artificial Intelligence Open Network (LAION)

Remote | 07/2025 - Present

- Worked on multiple open-source research projects, including CLAP (Contrastive Language Audio Pre-Training) and Project Alexandria, a large-scale effort to free research from copyright into a unified, comprehensive knowledge base.

Engineering Society of Detroit, Girls in Engineering

Dearborn, MI | 07/2023 - 08/2023

- Taught underrepresented, at-risk girls how to reason and program in Python for a summer. Increased test scores by 40%.

PROJECTS

Sheet Music Diffusion Generator - Carnegie Mellon University

Fall, 2024

- Created a diffusion network for creating sheet music with two related u-nets and VAE-disentanglement.

LLama-2 LLM Implementation - Carnegie Mellon University

Fall, 2024

- Implemented LLama-2 LLM with Rotary Positional Embeddings (RoPE) and Grouped Query Attention using PyTorch.

Text-to-image Generative Diffusion Network - Carnegie Mellon University

Fall 2024

- Implemented and trained a text-to-image model with cross-attention.

Sketch-to-image Generative Diffusion Network - Carnegie Mellon University

Spring, 2024

- Created a diffusion based network to create hand-drawn sketches from real photographs. Designed architecture and
- Designed architecture, composite loss function, and training procedure from extensive research and trial.

Host-Based Anomaly Intrusion Detection System - Carnegie Mellon University

Spring, 2024

- Created an intrusion detection system to malware from Linux process activity via. syscalls on AWS EC2 for 91% accuracy.

Other projects: *Numpy Neural Network Implementation, Radar Denoising GAN, C2 Prediction Web-App, Full-Featured Compiler.*

SKILLS

ML Engineering: *PyTorch, generative AI, CNNs, attention, transformers, diffusion, graph neural networks, data engineering.*

Cloud & Cloud ML: *AWS EC2 and Trainium, DynamoDB, PostgreSQL, PySpark, Kafka, Neo4j GraphDB, ML on the edge.*

Programming Languages: *Python, C/C++, BASH, GoLang, Java, JavaScript, Arm64/x86/MSP430 ASM, and more.*