Colin Wolfe

Washington, D.C. · (540) 920-8006 · colin.h.wolfe.27@dartmouth.edu
coho905.github.io · linkedin.com/in/colin-h-wolfe

Education

Dartmouth College, Hanover, NH — **Expected 2027**
Bachelors, Double Major in Computer Science and Mathematics, Minor in Public Policy (Major GPA: 4.0). — **GPA 3.81/4.0**
Relevant Coursework: Machine Learning, Linear Algebra, Algorithms, Intro to Public Policy, The Price System
Activities: Quant Club, Magnuson Center of Entrepreneurship, Phi Delta Alpha Fraternity, DCR Radio, *The Dartmouth Newspaper*

Commonwealth Governor's School, Spotsylvania, VA — **May 2023**
Honors/Awards: Valedictorian, Dunkin Donuts' DMV Scholar, National Rural Scholar, SAT Score: 1550/1600 — **GPA 4.0/4.0**

Work Experience

Machine Learning & Network Science Lab — *Researcher* — **Jan 2025-Present**
- Conduct advanced research on graph neural networks (GNNs) under the guidance of Professor Yan, focusing on applications in Large Language Models and Knowledge Maps. Teaching LLM agents to debate and rank their arguments.

Benchify — *Machine Learning Engineer* — **Jan 2025-Present**
- Leading the development of AI-driven code analysis tools at a Y Combinator backed startup, focusing on function context extraction, Abstract Syntax Tree (AST) parsing, and deeper program comprehension to enhance software workflows.
- Designing and optimizing automated code review and testing systems, leveraging machine learning to improve code intelligence, detect subtle bugs, enforce best practices, and enhance developer productivity and software reliability.

Distributed Information and Intelligence Analysis Group — *Researcher* — **Dec 2024-Present**
- Conducting advanced research in nonlinear decision-making, emergent learning, and multi-source fusion with a focus on innovative reasoning for Large Language Models. Nominee for the Presidential Scholars and Stamps Scholarship.

Versara.ai — *Co-Founder* — **Sep 2024-Present**
- Founded a startup based on protecting intellectual property from AI scrapers and tools like Perplexity. Raised over $10,000 in pre-seed funding. Successfully built a working prototype that blocks the most capable web scrapers available.
- In talks to provide protections to a blogging service with 500,000 users and a major global media organization with over 5 million digital subscriptions. Created a novel data poisoning algorithm, scalable servers, and the website www.versara.ai.

Probity Inc. — *Software Engineering Intern* — **May 2024-Aug 2024**
- Developed a new human-in-the-loop machine learning paradigm for Spoken Language Verification called Online Active Learning with Corrective Feedback. Achieved results two hundred times better than traditional training.
- Ran extensive experiments, configured and deployed additional servers for computational tasks on Intel NUCs, optimized machine learning algorithms for CPU performance, and created a corpus exceeding 100 GB of South Asian languages.

DALI Lab — *Machine Learning Engineer and Project Lead.* — **Nov 2023-Dec 2024**
- Developed and implemented advanced machine learning algorithms and data preprocessing methods to enable accurate image recognition of plastic symbols. Became a Neukom Scholar for the development of novel computational techniques.
- Led teams of machine learning engineers on several projects, including PlastiCycle and the National Park Service's BarnacleVision. Perform code review, technical & team leadership, mentoring, conducting design plans, and more.

Thomas Jefferson National Particle Accelerator Facility — *Engineering Intern* — **Jun 2022-Jul 2022**
- Created AI-based surrogate models of scientific code for the PHASM project (Parallel Hardware viA Surrogate Models).
- Built and implemented advanced physics-informed neural networks to approximate differential equation solutions related to accelerator experiments. Created bash scripts to streamline the installation process into fewer steps on various OS's.

Personal Projects

Predictive Market Analysis Program — **Jun 2022-Sept 2022**
- Built a program to predict Dow Jones market trends, using weather data near the NYSE, with >85% accuracy. Based on a theoretical literature review I wrote. Implemented in C++, used statistical methods & machine learning on 2 years of data

Tiny Search Engine — **Jan 2024-Mar 2024**
- Developed a crawler, indexer, and querier to return search results in the https://cs50tse.cs.dartmouth.edu/tse/ database. Implements a page-rank algorithm to match advanced queries involving logical conjunctions. Built solely in C.

Skin Cancer Classifier — **Aug 2021-Jun 2023**
- Built a high-accuracy convolutional neural network using Fast.ai to classify images of skin as cancerous or not. Built a website to handle uploaded pictures and showcase their predictions and associated probabilities. Used Python and Flask.

Thale Programming Language — **Nov 2024-Dec 2024**
- Building a bytecode interpreter for my programming language, Thale. Thale has user input capabilities, classes, first-class functions, inheritance and other OOP principles, variables, file execution or interactive prompts, garbage collection

Other Projects — **Aug 2019-Present**
- Speech based shell interface, Part of speech tagger, Rust version of the wc bash command, Zero-loss file (de)compressor

Skills, Certifications, & Interests

Programming Languages: Python, Rust, SQL, C, C++, R, Java, Bash, HTML/CSS, JavaScript
Technologies: AWS, Git, Docker, React, Scikit-Learn, PyTorch, TensorFlow, Fast.ai, Jupyter Notebook, Deep Learning, Flask
Interests: Financial Markets, Learning Spanish, Livestock Farming and Showmanship, Camping, Grand Strategy Games