Colin Wolfe

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

Education

Dartmouth College, Hanover, NH — **Expected 2027**
Bachelors in Computer Science and Mathematics, Minor in Public Policy — **Major GPA 4.0/4.0**
Relevant Coursework: Machine Learning, Algorithms, Deep Learning, Robotics, Linear Algebra, Secure Software Design
Commonwealth Governor's School, Spotsylvania, VA — **May 2023**
Honors/Awards: Valedictorian, Dunkin Donuts' DMV Scholar, National Rural Scholar — **GPA 4.0/4.0**

Work Experience

Benchify — *Machine Learning Engineer* — **Jan 2025 - Mar 2025**
- Developing AI-driven code analysis tools at a Y Combinator-backed startup, improving automation workflows with ML.
- Refining Retrieval-Augmented Generation (RAG) pipelines to strengthen code suggestions, deepen contextual function understanding, and generate explanations by utilizing embeddings, similarity search, and other retrieval techniques.

Distributed Information and Intelligence Analysis Group — *Researcher* — **Dec 2024 - Present**
- Researching probabilistic evaluation frameworks and structured training strategies to analyze emergent learning, quantify capability accumulation, and examine the factors that drive the development of complex reasoning.
- Optimizing architecture and data-efficient training by integrating curriculum learning, knowledge distillation, and multi-source fusion to enhance generalization in small models. Nominated for the Stamps and Presidential Scholars Programs.

Versara.ai — *Co-Founder* — **Sep 2024 - Present**
- Co-founded a startup securing digital content and from AI scrapers, raising $10,000+ in pre-seed funding.
- Negotiating deals with a service that has 500,000+ users and a media organization with over 5 million digital subscribers.
- Built a novel data poisoning algorithm, scalable servers, and a robust web platform that successfully blocks all known AI scrapers, including Perplexity and other advanced extraction tools, ensuring comprehensive content protection.
- Presented at MIT Privacy and Security Night on content scraping and demonstrated Versara's data poisoning approach.

DALI Lab — *Machine Learning Engineer and Project Lead.* — **Nov 2023 - Present**
- Lead teams of machine learning engineers on multiple projects, including the National Park Service's BarnacleVision.
- Conducting technical execution, performing code reviews, mentoring engineers, and designing project architectures.
- Implement and develop advanced machine learning algorithms and data preprocessing techniques for image recognition, predictive modeling, deployment. Awarded Neukom Scholarship for the development of novel computational methods.

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.

Thomas Jefferson National Particle Accelerator Facility — *Software 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

Speech to Terminal (SPOT) — **Nov 2024 - Dec 2024**
- Built a voice-controlled natural-language shell interface in Python, integrating speech recognition for hands-free terminal navigation. Integrated open-source voice-to-text and OpenAI LLM API for seamless text-to-command execution.

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. Implemented in C.

Skin Cancer Classifier — **Sep 2022 - Jun 2023**
- Developed a high-accuracy convolutional neural network with Fast.ai to determine the potential malignancy of skin anomalies. Designed a web platform for seamless image uploads and real-time predictions using Python and Flask.

Predictive Market Analysis Program — **Jun 2022 - Sept 2022**
- Created a program to predict Dow Jones market trends, using weather data near the NYSE, with >85% accuracy. Based on a literature review I wrote. Applied statistical models and machine learning to market and weather data in C++.

Other Projects — **Aug 2019 - Present**
- Part of speech tagger, Rust version of the wc bash command, Mammogram mass analyzer, Zero-loss file (de)compressor.

Skills, Interests

Programming Languages: Python, Java, C, C++, SQL, Rust, Bash, HTML, CSS, JavaScript
Skills: AWS, Git, Docker, Scikit-Learn, PyTorch, TensorFlow, Fast.ai, Jupyter Notebook, Flask, React, LangChain, OLlama, Pandas, Numpy, Nginx, OpenCV, FastAPI, Valgrind, GDB, *nix Systems, Mariadb, Wireshark, Technical Writing, Leadership
Interests: Artificial Intelligence, Financial Markets, Philosophy, Robotics, HCI, Reliable Software, AI Agents, Writing, Hiking
Activities: Dartmouth Entrepreneurship Club, Phi Delta Alpha Fraternity, Dartmouth Political Union, Writer for *The Dartmouth*, Dartmouth Leadership Attributes and Behaviors Program, Editor on *Dartmouth Law Review,* WebDCR Radio Show Host