Thomas Cummings | cummings.t287@gmail.com | www.linkedin.com/in/tom-r-cummings

## Skills

- **Languages**: Python, TypeScript, Rust, Elixir
- **Tools:** Git, Linux CLI, Docker, Make (and CMake), Terraform, PyTest and Jest
- **Infrastructure**: AWS (EC2, ECS, Lambda, S3)

## Experience

**Senior Software Engineer,** Smile ID — July 2025 - Present

- Feature development on the company's identity verification and antifraud systems, with a focus on liveness and user deduplication.
- Rebuilt the company's embeddings database; cut the operational cost of the system by ~40% while serving millions of inserts and searches a day with a p99 of < 200 ms.
- Maintain an inference service composed of a suite of machine learning models and LLMs in a production environment using the Triton framework and **Kubernetes**.

**Data Engineer, Vision Lead**, Reality Defender — December 2023 - July 2025

- Designed high-throughput image and video preprocessing pipelines based on well known computer vision algorithms for ingestion, augmentation, and feature extraction using **Python**, **FFmpeg**, **C++**, and **Docker**, deployed to **AWS ECS** reproducibly with **Terraform**.
- Developed a distributed data streaming service written in **Elixir** which replaced block-storage based data delivery for model training, saving tens of thousands of dollars in storage costs while maintaining model iteration speed.
- Rewrote the research team's **PyTorch** data-loader to increase GPU utilization and implement the **gRPC** interface used by the data streamer, using **Python**'s multiprocessing primitives.
- Built unit and integration test suite for data pipelines using **Pytest**.

**Application Developer**, ICF International — October 2022 – December 2023

- Programmed computer-based interview systems for 3 country-wide household surveys in developing countries using **JavaScript** and **HTML**; administered remote data collection systems for low-connectivity environments using low-payload apps with on-device storage and custom server implementations**.**
- Developed data analysis pipelines for large-scale surveys, going from raw data to publication-ready tables using **Python** and GIS libraries
- Automated **QA** and **validation** tools and test suites to ensure real-time data integrity in remote deployments, minimizing error propagation in downstream systems.
- Provided technical training and documentation to international field teams, scaling adoption of digital data systems and improving operational efficiency.

**Technical Systems Advisor**, The Carter Center — July 2022 – October 2022, Contracted

- Designed geospatial data analysis workflows using **Python** and **QGIS** to model disease transmission patterns across seasonal water sources, laborer migration, and village networks.
- Built a web dashboard in **JavaScript** to prioritize intervention zones based on terrain accessibility, human movement data, and historical infection reports.
- Supported distributed operations by coordinating real-time data collection across 150+ rural volunteers, developing tooling and protocols to improve surveillance reliability and streamline reporting under intermittent network conditions.

**Malaria Team Fellow**, Population Services International — Summer 2021, Internship

- Analyzed 3 years of survey data from across 300 households in rural Kenya to determine the average lifespan of malaria nets distributed; used regression modeling in Python to determine what household factors contribute to net failure
- Led systematic literature review of >300 peer-reviewed articles to generate context specific recommendations for improving integrated community case management practice

**Data Engineer**, Bowery Residents' Committee February 2020 – December 2020

- Planned and executed a full database migration to **MySQL**, modernizing backend infrastructure and enabling scalable reporting for homelessness services across New York City.
- Developed and maintained reporting pipelines delivering four monthly operational dashboards tracking outreach performance, client services, and placement trends, consumed by NYC Department of Homelessness Services.

**Technical Field Specialist**, Peace Corps & Last Mile Health June 2017 – December 2019

- Built a custom XML validation tool in **Python** aimed at early detection of transfer issues coming from payloads generated by the Open Data Kit application, later contributing bug fixes to the project.
- Designed and maintained a spatial database and visualization suite using **MySQL**, QGIS, and **Python**, modeling access gaps across villages and seasonal routes to inform health intervention logistics.
- Led technical operations for field **QA** efforts, supporting community health workers and coordinating with local teams to improve service data integrity in low-connectivity environments.

## Projects

### Chip-8 Interpreter

https://github.com/TomRCummings/Chipulator

- Created a Chip-8 interpreter in **C++**, including opcode parsing, stack/**virtual machine** management, and I/O handling.
- Included support for saving and loading interpreter state, step-through execution, and live viewing of the virtual machine state at runtime.
- Built screen and sound components using **SDL2** library, integrating GUI elements with **WxWidgets**

### Gopher Client

https://github.com/TomRCummings/Zero

- Created a cross-platform, terminal-based Gopher client, implementing RFC 1436, using modern **C++**
- Added enthusiastic fetching of selectors and caching based on hostname/selector pairs, cutting down on re-requests during the life of user session
- Built a light-weight abstraction layer over **Linux** and Windows **socket APIs**

### Game Boy Bare Metal "Hello World"

https://ninimotom.com/writing/gb-hello-world

- Demonstrated how to print text to a Nintendo Game Boy screen with hand-written **assembly**, **linker scripts**, and a custom font.
- Walked through the Game Boy's memory architecture, including RAM, VRAM, and I/O registers, to load a font, print to the screen, and wrap lines that exceed the screen space of the machine.
- Encoded font glyphs into the 2bpp tile graphics format required by the Game Boy's PPU and implemented routines for loading tile data and writing to the tilemap.
- Wrote a detailed technical blog post with working source code, annotated assembly, and linker directives explaining each part of the process.

## Education

**Master of Public Health, Epidemiology (4.0 GPA)**
Columbia University Mailman School of Public Health
Thesis: *Designing a machine learning pipeline with a fine-tuned BERT for measuring associations between COVID-19 and long COVID symptoms using self-reported syndromic information from a social media cohort*

**Bachelor of Science, Biochemistry**
The Ohio State University
Relevant Coursework: CSE 2122 (Data Structures Using C++), CSE 2221 (Software Components), CSE 2231 (Software Design), CSE 2321 (Discrete Structures), CSE 2331 (Data Structures and Algorithms)