Improved Situational Awareness through AI at the Edge

PHITE: Portable, High-Performance Inference at the Tactical Edge

ON BATTLEFIELDS AND IN OTHER HARSH OR CONTESTED ENVIRONMENTS, TACTICAL EDGE DEVICES, SUCH AS SENSORS, WEARABLE TECHNOLOGIES, AND ULTRA-LIGHTWEIGHT UAVS, are key to achieving decision advantage. The ability to process and analyze data where it is collected allows the military to make decisions more quickly and within an opponent's observe-orient-decide-act (OODA) loop. While the ability to view what is happening around a corner, over a ridge, or above a building provides our forces with increased situational awareness from afar, what happens when our forces need to augment those views with intelligence? Many tactical edge devices lack the memory and power to process machine learning (ML) workloads that can provide relevant intelligence.

Our research project, Portable High-performance Inference at the Tactical Edge (PHITE), aims to address the challenges of performance engineering, embedded systems, and AI to bring ML to these edge devices, giving our military forces greater situational awareness and laying foundations for scalable AI that can work anywhere.

Supporting DoD AI and ML Workloads on Ultra-Low Power Devices

To move sophisticated analytics capabilities onto edge devices, we must meet challenges in the fields of performance engineering, embedded systems, and AI, specifically ML.

Most performance engineering research to date has focused on dense linear algebra rather than the sparse, irregular computations for ML workloads—and on mainstream CPUs and GPUs rather than small embedded devices. Similarly, most ML re-search in this area has focused on shrinking ML models developed on larger platforms rather than improving the computational frameworks that use the models already intended for embedded systems.

A Library of High-Performance, Low-Power ML Algorithms

To meet the challenge of ML on edge devices, we will develop an open source, portable library of ML algorithms for embedded devices.

After first selecting and characterizing baseline applications and hardware platforms, we will demonstrate the effectiveness of the algorithms in a lab environment and then in representative environments. Our aim is to develop the library in a way that is more easily optimized for target platforms and applications, with a goal of achieving a 1.5x-10x improvement in performance while maintaining accuracy. As members of the TinyML community, we will help guide the development of relevant benchmarks and use the benchmark reference implementations to evaluate our own work.

Improvements in Power Efficiency and Data Processing Capabilities

PHITE will provide DoD organizations with the ability to deploy sophisticated analytics on edge devices—with a possible 10x or more improvement in power efficiency for machine-learned model inference capabilities. The impact on capability could include processing more data collected by high-fidelity sensors; increasing in precision identification of objects, activities, or patterns of interest; or extending the lifetime of a battery-powered smart sensor. Providing the ability to package sophisticated data analytics capabilities in smaller computational packages could also provide significant cost savings. Finally, the results of this research will support the timely development of new smart edge-sensor capabilities or the migration of existing edge analytics to new computing devices through software libraries designed for usability and cross-platform portability with a reduction in time to delivery.

AI Engineering: Scalable AI

As we work to meet the challenges of moving ML to ultra-low-power edge devices, we will make foundational contributions to AI engineering, an emergent discipline focused on developing tools, systems, and processes to enable the application of artificial intelligence in real-world contexts. PHITE will yield important insights into creating scalable AI: algorithms, data, models, and computing infrastructure that accommodate the size, speed, and complexity requirements of mission needs.

Areas of Opportunity

PHITE shows promise for advancing ML on the edge for

- Force protection
- Predictive maintenance
- Persistent surveillance
- Humanitarian operations
- Autonomous military robotics
- Space operations and intelligence

Work with Us

Principal Investigator
DR. SCOTT MCMILLAN
smcmillan@sei.cmu.edu

Program Manager
FRANK REDNER
fredner@sei.cmu.edu

About the SEI

The Software Engineering Institute is a federally funded research and development center (FFRDC) that works with defense and government organizations, industry, and academia to advance the state of the art in software engineering and cybersecurity to benefit the public interest. Part of Carnegie Mellon University, the SEI is a national resource in pioneering emerging technologies, cybersecurity, software acquisition, and software lifecycle assurance.

Contact Us

CARNEGIE MELLON UNIVERSITY
SOFTWARE ENGINEERING INSTITUTE
4500 FIFTH AVENUE; PITTSBURGH, PA 15213-2612

sei.cmu.edu
412.268.5800 | 888.201.4479
info@sei.cmu.edu