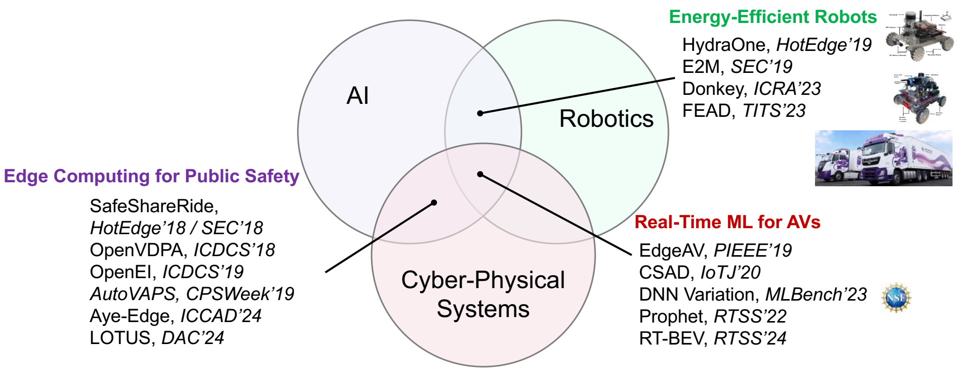
Research Statement

Liangkai Liu, University of Michigan

Modern cyber-physical systems (CPS) are rapidly advancing toward autonomy, creating extraordinary opportunities alongside demanding real-time and safety challenges. Autonomous driving is a clear example—redefining mobility yet requiring strict guarantees on reliability, timing, and trustworthiness.

> My research aims to build **predictable** and **efficient** computing systems for AI-enabled, safety-critical CPS—spanning autonomous vehicles, intelligent robots, and edge-connected public safety systems.

My work addresses the tension between AI's computational demands and the physical constraints of real-time systems by rethinking the end-to-end software stack—from algorithms to middleware/RTOS to heterogeneous hardware. For autonomous driving, I develop *real-time ML* that couples perception/planning models with runtime scheduling to deliver deadline-aware inference under uncertainty. For robotics, I design *energy-efficient autonomy pipelines* that co-optimize model architectures, execution paths, and accelerator mappings to meet power budgets without sacrificing safety. For public safety, I build *edge-computing architectures* that distribute sensing and decision-making across vehicles, infrastructure, and the cloud, enabling time-critical responses in dynamic environments.



These efforts are unified by a cross-layer perspective that (i) *co-optimizes* algorithms, middleware (e.g., ROS/ROS2), OS (e.g., Real-Time Linux, RTOS), and hardware (GPU/CPU) to reduce end-to-end jitter; (ii) uses *profiling-driven optimization* to expose true bottlenecks across compute, I/O, memory, and contention; and (iii) applies *performance–time tradeoffs*—such as early exits, ROI-aware processing, and precision/activation scaling—to meet real-time bounds while saving energy. The result shows it not only works in theory but *performs reliably and responsibly in the real world*.

Research Highlights: My research is inherently cross-disciplinary—bridging CPS, AI/ML, real-time systems, robotics, intelligent transportation, and edge computing. I have published in top venues such as RTSS, RTAS, ICCAD, DAC, ICRA, SEC, HotEdge, TITS, IWC, and IoTJ. I was among the first in the computing-systems community to investigate DNN *inference-time variability* in autonomous driving, leading to a **$600,000 NSF grant** as a **Co-Principal Investigator** on predictable multi-tenant DNN inference. I developed testbeds (HydraOne, Donkey) for programmable, energy-efficient autonomous mobile robots, and my fuel-efficient autonomous trucking research achieved a **5%** fuel reduction in practice compared to experienced drivers. As of October 2025, my work has received over **2,500** citations, with an H-index of **19** and an i10-index of **22**.

Prior Research

1. Real-Time ML for AVs

Ensuring predictability of deep neural networks (DNNs) in AV's perception pipeline presents significant challenges [1]. To address this issue, we conducted a comprehensive profiling of DNN inference time variations [2] and developed novel solutions to achieve predictable perception for AVs. First, `Prophet` mitigates substantial time variations in most DNN models within an AV system through novel early-exit and coordination mechanisms [3]. Second, `RT-BEV` addresses multi-camera synchronization and detection issues through a co-design of communication and detection, coupled with ROI-based processing [4]. Finally, `AyE-Edge` presents a general solution for power-efficient real-time object detection via the co-design of keyframe selection, model pruning, and CPU-GPU configuration [5].

Inference Time Variations. We developed a reconfigurable testbed for DNN inference profiling [6]. On top of that, we analyze the time variation in DNN inference in fine granularity from *six perspectives: data, I/O, model, runtime, hardware, and end-to-end perception system*. Six insights are derived in understanding the time variations for DNN inference [2]. This work contributes to two research papers and one **NSF Grant** for studying *DNN inference time variations for multi-tenant DNNs*.

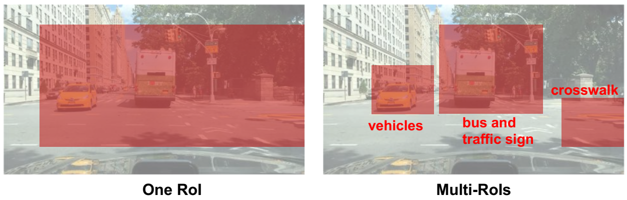


Figure 1: An example of environment-aware dynamic ROIs.

Mitigating Inference Variations. We identified root causes for DNN inference time variations in AVs for the *first* time through in-depth profiling. We found that the variations in inference time for a single DNN model are mainly caused by the *DNN's multi-stage/multi-branch structure*, which has a dynamic number of proposals or raw points. In addition, we found that *uncoordinated contention and cooperation* are the root causes of the time variations in multi-tenant DNN inference. Based on these insights, we proposed the `Prophet` system that addresses the time variations in the AV perception system in two steps [3]. The first step is to predict the time variations based on the intermediate results, like proposals and raw points. The second step is coordinating the multi-tenant DNNs to ensure the execution progress is close to each other. The evaluation on the KITTI dataset shows effective performance in reducing inference time variations.

ROI-Aware BEV Perception. We proposed `RT-BEV`, the *first* framework designed to *co-optimize message communication and object detection* to improve real-time e2e BEV perception without sacrificing accuracy [4]. The main insight of `RT-BEV` lies in generating *traffic environment- and context-aware Regions of Interest (ROIs)* for AV safety, combined with ROI-aware message communication. `RT-BEV` features an ROI-aware Camera Synchronizer that adaptively determines message groups and allowable delays based on ROIs' coverage. We also develop an ROIs Generator to model context-aware ROIs and a Feature Split & Merge component to handle variable-sized ROIs effectively. Furthermore, a Time Predictor forecasts timelines for processing ROIs, and a Coordinator jointly optimizes latency and accuracy for the entire e2e pipeline. We have implemented `RT-BEV` in a ROS-based BEV perception pipeline and evaluated it with the nuScenes dataset. `RT-BEV` significantly enhances real-time BEV perception.

Power-Efficient Real-Time Object Detection. We designed `AyE-Edge`, the *first-of-this-kind* development tool that explores automated algorithm-device deployment space search to realize accurate yet power-efficient real-time object detection on the edge [5]. Through a *collaborative exploration of keyframe selection, CPU-GPU configuration, and DNN pruning strategy*, `AyE-Edge` excels in extensive real-world experiments conducted on a mobile device. The results consistently demonstrate `AyE-Edge`'s effectiveness, realizing outstanding real-time performance, detection accuracy, and notably, a remarkable 96.7% reduction in power consumption, compared to state-of-the-art (SOTA) competitors.

2. Energy-Efficient Autonomy

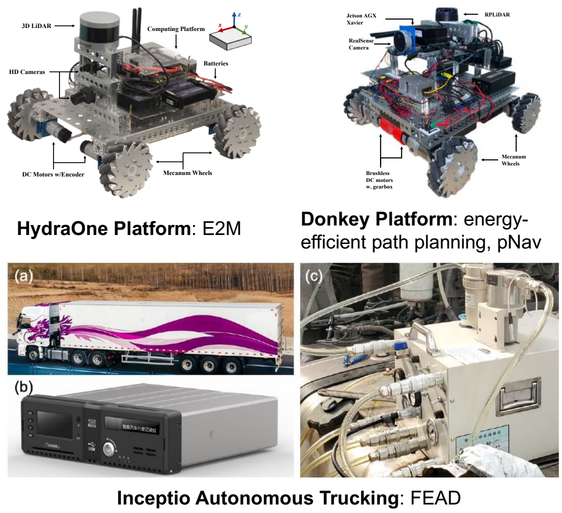


Figure 2: Testbeds for Energy Efficient Autonomous System.

Energy efficiency has been a fundamental issue for autonomous mobile robots and heavy-duty trucks. Autonomous driving techniques bring potential for saving energy through proper coordination of the software stack (perception, planning, and control) and hardware platforms (sensor, computing device, motors, engine management system). Figure 2 shows platforms we built for energy efficinet autonomous system. More specifically, on top of the practical autonomous mobile robots (AMRs) like HydraOne platform [7], E2M coordinates the perception module with computing devices and sensors [8]; Donkey platform extends this to the whole navigation stack with sensor, computing device, and motors [9]; FEAD coordinates planning and control with the truck's engine management systems [10]; 4C [11] represents a visinary design for co-design computation, communication and control for end-to-end efficiency.

Energy-Efficient Computation Middleware. We discovered three main *root causes* of energy inefficiency for AMR's computing platform: *uncoordinated access to sensor data*, *performance-oriented model inference execution*, and *uncoordinated execution of concurrent jobs*. To address three inefficiencies, we proposed E2M, the *first* energy-efficient middleware software stack for autonomous mobile robots [8]. First, E2M regulates the access of different processes to sensor data, e.g., camera frames, so that the amount of data actually captured by concurrently executing jobs can be minimized. Second, based on a predefined per-process performance metric (e.g., safety, accuracy) and desired target, E2M manipulates the process execution period to find the best energy-performance trade-off. Third, E2M coordinates the execution of the concurrent processes to maximize the total contiguous sleep time of the computing hardware for maximized energy savings. Our experimental results on the HydraOne platform show that E2M leads to 24% energy savings for the computing platform, which translates into an extra 11.5% of battery time and 14 extra minutes of robot runtime, with a performance degradation lower than 7.9% for safety and 1.84% for accuracy.

Energy-Efficient AMRs. We identified a lack of coordination between computation and control as a major source of energy inefficiency for the whole AMRs platform. Therefore, we proposed a *comprehensive* energy prediction model that provides real-time energy consumption for each component of the AMR [9]. Additionally, we propose three path models to address the obstacle avoidance problem for AMRs. To evaluate the performance of our energy prediction and path models, we have developed a customized AMR called Donkey, which has the capability for fine-grained (millisecond-level) end-to-end power profiling. Our energy prediction model demonstrated an accuracy of over 90% in our evaluations. Finally, we applied our energy prediction model to obstacle avoidance and guided energy-efficient path selection, resulting in up to a 44.8% reduction in energy consumption compared to the baseline.

Fuel Efficient Autonomous Trucking. Fuel cost contributes significantly to the high operation cost of heavy-duty trucks. State-of-the-art models show poor performance and are rarely deployed in practice, limited by accurate features directly related to the truck's fuel consumption. We used the *truck's engine management system (EMS) and Instant Fuel Meter (IFM)* to collect a three-month dataset during the period of December 2019 to June 2020 [10]. We designed seven prediction models, including linear regression, polynomial regression, MLP, CNN, LSTM, CNN-LSTM, and AutoML, are investigated and evaluated them to predict real-time fuel rate. The evaluation results show that the EMS and IFM datasets help to improve the coefficient of determination of traditional linear/polynomial models from 0.87 to 0.96, while the learning-based approach, AutoML, improves the coefficient of determination to attain 0.99. Besides, the actual deployment of fuel rate prediction and path planning shows up to 5% fuel saving on real Inceptio trucks [12].

Communication, computation, and control Co-design. To improve the end-to-end performance, we envision that future CAVs require co-design of communication, computation, and control [11]. We designed 4C, which provides a unified communication, computation, and control co-design framework. With *programmable communications, fine-grained heterogeneous computation, and efficient vehicle controls* in 4C, CAVs can handle critical scenarios and achieve energy-efficient autonomous driving.

3. Edge Computing for Public Safety

Connected and Autonomous Vehicles (CAVs) are becoming more and more popular for the automobile academic and industry community. Edge computing has become a promising solution to enhance the safety of CAVs [13]. I proposed SafeShareRide, which is an edge computing-enabled system for protecting both driver and passenger's safety [14, 15]. We also designed AutoVAPS, which is an edge-based system that detects safety events for law enforcement officers in real-time. Furthermore, I have worked on building an open and comprehensive vehicular data analytics platform to support a variety of applications from AD/ADAS to real-time diagnostics, driver behavior detection, and infotainment.

Ridesharing Safety. Ridesharing services, such as Uber and Didi, are enjoying great popularity; however, a big challenge remains in guaranteeing the safety of passengers and drivers. We propose an edge-based attack detection in ridesharing services, namely `SafeShareRide`, which can *detect dangerous events happening in the vehicle in near real-time* [15, 14]. `SafeShareRide` is implemented on both drivers' and passengers' smartphones. `SafeShareRide` is the world's first framework that can support real-time attack detection in ridesharing services.

Law Enforcement Officer/Vehicle Safety. We proposed an IoT-Enabled public safety service called *AutoVAPS* [16] which integrates body-worn cameras and other sensors on the vehicle for public safety. In *AutoVAPS*, we propose a reference architecture that consists of the data layer for data management, the model layer for edge intelligence, and the access layer for privacy-preserving data sharing and access. Object detection is implemented as a case study of *AutoVAPS*.

Open Data Analytics Platform. we designed OpenVDAP [17] which is a full-stack edge based platform including an on-board computing/communication unit, an isolation-supported and security & privacy-preserved vehicle operation system, an edge-aware application library, as well as an optimal workload offloading and scheduling strategy, allowing CAVs to dynamically detect each service's status, computation overhead and the optimal offloading destination so that each service could be finished within an acceptable latency and limited bandwidth consumption. Most importantly, in contrast to the proprietary platform, OpenVDAP is an open-source platform that offers free APIs and real-world vehicle data to the researchers and developers in the community, allowing them to deploy and evaluate applications in a real environment.

Future Directions

Looking ahead, my research will address critical challenges in *real-time, safe, and secure cyber-physical systems and intelligent transportation*. By advancing methodologies in these domains, I aim to enhance the robustness, reliability, and performance of such systems in complex and dynamic environments.

End-to-End CAVs Safety. I plan to develop comprehensive *end-to-end* profiling methodologies for connected and autonomous vehicles (CAVs) that cover every aspect of their operation—from perception to control. This holistic approach is essential for ensuring robustness and safety across a wide range of real-world scenarios. A key component of this work will be the creation of *universal metrics applicable to diverse security challenges*, enabling accurate assessment of different strategies' strengths and limitations. These metrics will evaluate multiple facets of the AV pipeline, helping to identify where attacks and defenses are most effective and to quantify their impact on functional safety. By providing deeper insights into *systemic threats and vulnerabilities*, this research will contribute to the development of more secure and reliable autonomous vehicle systems.

CAVs Security. To improve the safety and trustworthiness of CAVs, I plan to investigate *vulnerabilities that span the entire system stack*—from hardware and sensing to perception, planning, and control. Rather than treating each layer in isolation, this research will adopt a *holistic, cross-layer approach* to understand how failures or attacks in one module can propagate and compromise overall system integrity. I will study diverse attack vectors—such as sensor spoofing, adversarial ML inputs, and control hijacking—and examine how they interact across component boundaries. By integrating insights from hardware faults, software vulnerabilities, and AI model behavior, I aim to develop *coordinated defense strategies* that address threats at multiple abstraction levels. This work will help establish a unified framework for assessing and enhancing the *security posture of general CAV systems*, ultimately contributing to the development of more resilient and trustworthy autonomous platforms.

Predictable Offroad Autonomous Driving. I also aim to tackle the unique challenges of offroad autonomous driving, where *unpredictable conditions*—such as fast-moving objects, irregular terrain, and sudden elevation changes—impose stringent, dynamic timing and accuracy requirements on ML-based perception pipelines. To address these challenges, I envision a *collaborative sensing architecture* that employs a heterogeneous sensor suite, including drones and advanced sensing technologies, to enhance situational awareness. By integrating diverse sensor inputs and developing robust perception and control algorithms capable of real-time adaptation, my goal is to design autonomous vehicles that can reliably and safely navigate offroad environments. This research will extend the reach of AV technologies to mission-critical domains such as agriculture, disaster response, and exploration.

Foundation Models in Cyber-Physical Systems. Foundation models in vision, language, and audio are increasingly being adopted in cyber-physical systems (CPS) for applications including autonomous driving, smart healthcare, and precision agriculture. These models empower CPS to perceive and reason about complex environments, enhancing decision-making and autonomy. However, integrating such computationally intensive models into resource-constrained platforms poses significant challenges—particularly with respect to energy efficiency and real-time performance. I advocate a *co-design approach* that jointly optimizes algorithms, system architectures, and hardware accelerators to enable efficient inference and training. This work will facilitate the practical deployment of foundation models in real-world CPS applications while maintaining high performance and low energy consumption.

References

[1] Liangkai Liu, Sidi Lu, Ren Zhong, Baofu Wu, Yongtao Yao, Qingyang Zhang, and Weisong Shi. Computing systems for autonomous driving: State of the art and challenges. IEEE Internet of Things Journal (IoTJ), 8(8):6469–6486, 2020.

[2] Liangkai Liu, Yanzhi Wang, and Weisong Shi. Understanding time variations of DNN inference in autonomous driving. MLSys Workshop on Benchmarking Machine Learning Workloads on Emerging Hardware, 2023.

[3] Liangkai Liu, Zheng Dong, Yanzhi Wang, and Weisong Shi. Prophet: Realizing a predictable real-time perception pipeline for autonomous vehicles. In IEEE Real-Time Systems Symposium (RTSS), pages 305–317. IEEE, 2022.

[4] Liangkai Liu, Jinkyu Lee, and Kang G. Shin. RT-BEV: Enhancing real-time BEV perception for autonomous vehicles. In 45th IEEE Real-Time Systems Symposium (RTSS), 2024.

[5] Chao Wu*, Yifan Gong*, Liangkai Liu*, Mengquan Li, Yushu Wu, Xuan Shen, Zhimin Li, Geng Yuan, Weisong Shi, and Yanzhi Wang. AyE-Edge: Automated deployment space search empowering accuracy yet efficient real-time object detection on the edge. In International Conference on Computer-Aided Design (ICCAD), 2024.

[6] Liangkai Liu, Yanzhi Wang, and Weisong Shi. CPT: A configurable predictability testbed for DNN inference in AVs. Tsinghua Science and Technology, 30(1):87–99, 2024.

[7] Yifan Wang, Liangkai Liu, Xingzhou Zhang, and Weisong Shi. HydraOne: An indoor experimental research and education platform for CAVs. In 2nd USENIX Workshop on Hot Topics in Edge Computing (HotEdge 19), 2019.

[8] Liangkai Liu, Jiamin Chen, Marco Brocanelli, and Weisong Shi. E2M: An energy-efficient middleware for computer vision applications on autonomous mobile robots. In Proceedings of the 4th ACM/IEEE Symposium on Edge Computing (SEC), pages 59–73, 2019.

[9] Liangkai Liu, Ren Zhong, Aaron Willcock, Nathan Fisher, and Weisong Shi. An open approach to energy-efficient autonomous mobile robots. In IEEE International Conference on Robotics and Automation (ICRA), 2023.

[10] Liangkai Liu, Wei Li, Dawei Wang, Yi Wu, Ruigang Yang, and Weisong Shi. Fuel rate prediction for heavy-duty trucks. IEEE Transactions on Intelligent Transportation Systems (TITS), 24(8):8222–8235, 2023.

[11] Liangkai Liu, Shaoshan Liu, and Weisong Shi. 4C: A computation, communication, and control co-design framework for CAVs. IEEE Wireless Communications, 28(4):42–48, 2021.

[12] Inceptio Completes Delivery of 300 Autonomous Heavy-Duty Trucks to YTO Express. https://en.inceptio.ai/news/99012.html.

[13] Shaoshan Liu, Liangkai Liu, Jie Tang, Bo Yu, Yifan Wang, and Weisong Shi. Edge computing for autonomous driving: Opportunities and challenges. Proceedings of the IEEE, 107(8):1697–1716, 2019.

[14] Liangkai Liu, Xingzhou Zhang, Mu Qiao, and Weisong Shi. Safeshareride: Edge-based attack detection in ridesharing services. In IEEE/ACM Symposium on Edge Computing (SEC), pages 17–29. IEEE, 2018.

[15] Liangkai Liu, Xingzhou Zhang, Mu Qiao, and Weisong Shi. SafeShareRide: Edge-based attack detection in ridesharing services. In USENIX Workshop on Hot Topics in Edge Computing (HotEdge 18), Boston, MA, 2018. USENIX Association.

[16] Liangkai Liu, Xingzhou Zhang, Qingyang Zhang, Andrew Weinert, Yifan Wang, and Weisong Shi. Autovaps: An iot-enabled public safety service on vehicles. In Proceedings of the fourth workshop on international science of smart city operations and platforms engineering, pages 41–47, 2019.

[17] Qingyang Zhang, Yifan Wang, Xingzhou Zhang, Liangkai Liu, Xiaopei Wu, Weisong Shi, and Hong Zhong. Openvdap: An open vehicular data analytics platform for cavs. In IEEE 38th International Conference on Distributed Computing Systems (ICDCS), pages 1310–1320. IEEE, 2018.