Achieving the Safety and Security of the End-to-End AV Pipeline

Noah T. Curran
ntcurran@umich.edu
University of Michigan
Ann Arbor, MI, USA

Minkyoung Cho
minkycho@umich.edu
University of Michigan
Ann Arbor, MI, USA

Ryan Feng
rtfeng@umich.edu
University of Michigan
Ann Arbor, MI, USA

Liangkai Liu
liangkai@umich.edu
University of Michigan
Ann Arbor, MI, USA

Brian Jay Tang
bjaytang@umich.edu
University of Michigan
Ann Arbor, MI, USA

Pedram MohajerAnsari
pmohaje@clemson.edu
Clemson University
Clemson, SC, USA

Alkim Domeke
adomeke@clemson.edu
Clemson University
Clemson, SC, USA

Mert D. Pesé
mpese@clemson.edu
Clemson University
Clemson, SC, USA

Kang G. Shin
kgshin@umich.edu
University of Michigan
Ann Arbor, MI, USA

Abstract

In the current landscape of autonomous vehicle (AV) safety and security research, there are multiple isolated problems being tackled by the community at large. Due to the lack of common evaluation criteria, several important research questions are at odds with one another. For instance, while much research has been conducted on physical attacks deceiving AV perception systems, there is often inadequate investigations on working defenses and on the downstream effects of safe vehicle control.

This paper provides a thorough description of the current state of AV safety and security research. We provide individual sections for the primary research questions that concern this research area, including AV surveillance, sensor system reliability, security of the AV stack, algorithmic robustness, and safe environment interaction. We wrap up the paper with a discussion of the issues that concern the interactions of these separate problems. At the conclusion of each section, we propose future research questions that still lack conclusive answers. This position article will serve as an entry point to novice and veteran researchers seeking to partake in this research domain.

CCS Concepts

• **Security and privacy** → **Embedded systems security**; • **Computer systems organization** → *Sensors and actuators*; *Robotics*.

Keywords

autonomous vehicle (AV), safety, security, cyber-physical system (CPS)

1 Introduction

As part of the proliferation of autonomous vehicles (AVs), automotive manufacturers have argued that they will improve the overall safety of roadways [30, 131]. For instance, after one-million driverless miles, Cruise has claimed that their "safety- minded" [29] self-driving technology is safer on average than a human driver [146]. Despite this, Cruise has faced immense scrutiny over their self-driving technology due to a series of recent accidents [58, 90, 116]. Some question whether a human driver would have caused these accidents [77]. In the aftermath, Cruise has recalled their AVs and ceased operation for the foreseeable future [31, 76].

Alongside the safety issues that AVs must still overcome, there have been advancements from academia of theoretical security concerns of AVs. These findings range from attacks on perception systems using adversarial ML techniques to tricking the object tracking systems. There are also more traditional attack vectors that use software exploits to gain unauthorized privilege escalation within the AV software stack. Coupling both the security and safety of an AV, assessing the robustness of the guarantees of correct AV operation is a complex procedure.

The structure of our position paper is as follows. First, we provide a brief background of concepts necessary for understanding AV safety and security. Then, we provide the current state of research from each identifiable community in the literature (summarized in Fig. 1). Each of these sections concludes with the limitations and takeaways.

In this article, we contribute the following:

- We provide a high-level overview of each major component to the AV. These components are summarized in Fig. 1.
- For each component of the AV, we present limitations of the current state-of-the-art research which has direct implications toward safety/security.
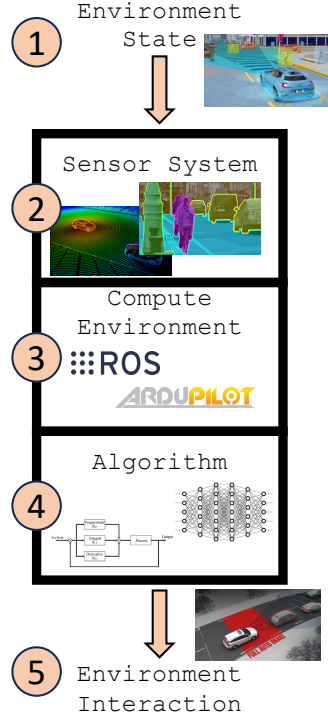
Figure 1: A summary of the components of the AV system. The focus is on the major topics of inquiry that concern the safety/security community. (1) Preserving the privacy of the environment around the vehicle. (2) Ensuring the correctness of sensor measurements. (3) Maintaining the integrity of the compute environment stack, including the middleware, RTOS, and CPU/GPU resources. (4) Proving algorithmic guarantees/robustness with respect to ensuring AV mission completion. (5) Providing safe and socially acceptable control of the AV as it interacts with its environment.

- To jump-start future research that may prioritize these limitations, we propose research questions for the major components of the AV, as well as the end-to-end system as a whole.

2 Background

In Fig. 1 we present a non-traditional overview of the components of the AV. Typically, most introductions to AVs present a *Sense-Plan-Act* control loop. However, *Sense-Plan-Act* does not have an abstraction that shows the important portions of the attack surface. In contrast, our representation combines the planning and control of the AV as part of the algorithmic processes, and we expose the middleware as a component between the sensor system and these algorithms. Our abstraction also raises importance to the environment before and after the AV is controlled. In this section we provide background necessary for understanding the goals of these components before we discuss the safety and security issues in the following sections.

2.1 Perception Sensors

In order to know what the world looks like around itself, the AV must use sensors that provide perception. The primary sensing units of the vehicle may include one or more camera, radar, LiDAR, or ultrasonic sensor [41]. For various other control or infotainment functions, the vehicle may also have a GNSS, gyroscope, accelerometer, microphone [65], or wireless transmitter/reciever [34]. On the same vehicle, there may be different forms of these sensors, such as thermal, monoscopic, or stereoscopic cameras; there are short-range and long-range radars; and LiDARs can use laser pulses, a continuous laser beam, a rotating laser beam, or solid state. The arrangement of the perception sensors will depend on the needs and requirements of the AV designer. The perception system may include a component for processing the data and extracting information critical to the operation of the vehicle. For instance, it may parse image data for important objects or segment it entirely into different critical regions. This data is then passed on to the safety-critical tasks that need it.

2.2 Middleware

Communication middleware is a foundational layer that facilitates efficient and reliable communication between the various software and hardware components of an AV. For instance, the various sensors are connected to the middleware, which manages and directs the sensor information toward the various tasks that require it. The middleware will also provide priority assignments to various tasks, as well as priviledge management for accessing information and overridding task assignment. The middleware also provides basic functionality for scheduling the tasks and meeting real-time requirements. These tasks may include algorithms that support safety or comfort features that may autonomously operate the vehicle. A few examples of a middleware include the Robot Operating System (ROS) [87, 105], Ardupilot [9], PX4 [104], Yet Another Robot Platform (YARP) [102], Cyber RT [12] and OpenRTM-aist [94].

2.3 AV Tasks

In the context of an AV, a task could be as simple as checking the distance of the vehicle to nearby objects, or as complex as a multi-object tracking algorithm. These tasks each have their expected execution time, deadline, and release time. Using this information, each task is assigned a priority so that the middleware knows which to execute next. Tasks that have safety-critical deadlines, those that have catastrophic consequences if missed, are called hard deadlines. Some deadlines may be missed without consequence, and are called soft deadlines. Those deadlines that do not lead to catastrophe if missed but cause the result of the task to be unusable are called firm deadlines. Depending on the task and environmental conditions, these deadline concepts may have some applicable variation, such as an (m,k)-firm deadline where at least m out of k instances of a task must complete execution before the deadline [54].

While the middleware can make assumptions about the tasks in order to quickly obtain a feasible schedule, they often have complex relationships with one another. For instance, some tasks may have higher priority based on the context (mixed-criticality system) [37], may not be executable until another dependent task completes [35], or may access a resource shared by another task [39]. AVs encounter variable and complex environments, so based on the context the

middleware may be required to enforce strict deadlines on specific tasks. Thus, flexibility of task priority assignment is necessary in order to maintain safe operation of the vehicle.

3 Environment State

For AVs, the environment state is crucial for vehicles' operation and interaction with their surroundings. The environment state refers to the comprehensive and real-time representation of all relevant factors and conditions surrounding the vehicle. This includes static elements such as road layouts, traffic signs, barriers, and driving setting as well as more dynamic elements such as current vehicle position, other vehicles, pedestrians, and weather conditions. An AV must accurately perceive and interpret the environment state as every action the AV takes, from changing lanes to adjusting speed, relies on an accurate representation of its current environment. For instance, the AV must be able to recognize a pedestrian crossing the street, understand that it signifies a need to slow down or stop, and take appropriate action.

3.1 Driving Environments

AVs drive through a vast number of public and private spaces, requiring them to navigate serious safety, security, and privacy concerns. At near-zero error rates, a level 4/5 AV is expected to handle complex contexts like urban landscapes and harsh weather. This leads to a large operational design domain, and AV companies utilize a find-and-fix approach when handling novel crash scenarios. The preference of this approach is due to trends in deep-learning approaches to use big data to improve the performance of models; the developers of the AVs are incentivized to maximize their data collection activities to improve training and evaluation sets with respect to minimizing accident statistics. To this end, AVs are equipped with information-rich and diverse sensors, making them capable of collecting incredibly sensitive data, particularly if driving scenes from multiple vehicles are collated.

3.2 Surveillance

Such sensor networks can be misused to perform targeted and/or mass surveillance [49] of individuals or communities. Already, car companies collect and sell copious amounts of sensitive and personal data with 3rd parties with little to no data controls or opt-outs for users [42]. In some instances, law enforcement and government organizations have requested self-driving car data without warrants [84]. This may be especially concerning if AV data is collected and aggregated from multiple vehicles. AV systems may become an omnipresent platform for collecting and inferring people's activities, a notion that many people expressed overwhelming discomfort about in a study by Bloom *et al.* [15]. Continuously monitoring people for the purposes of law enforcement or advertising seriously threatens personal autonomy and creates technological platforms ripe for the abuse of power [78].

3.3 Limitations & Takeaways

3.3.1 Existing Privacy Research. In the domain of AVs, there has been a serious lack of works discussing potential privacy risks and protection methods. Most AV privacy literature present legal arguments and concerns. While there is a heavy emphasis with research

on location privacy and biometric (face) privacy, little research has investigated privacy solutions in vehicles (e.g., opting out, privacy-aware data collection and sensing, etc.).

3.3.2 Proprietary ML Attacks. Xie *et al.* propose categorizing privacy risks into individual (people, attributes), population (communities, towns, demographics), and proprietary (ML models, architectures, sensor details), creating a taxonomy of attacks and defenses [138]. They, like most other works, tend to focus on adversarial ML attacks, (e.g., membership inference), differential privacy, or federated learning. [48, 55, 66, 103].

3.3.3 Sensor Privacy. We argue that this focus on proprietary privacy in AVs is insufficient, impractical, and fails to address the problem at its root cause – intrusive data collection and processing practices. Rather, privacy protection mechanisms should be integrated earlier in AV pipelines, e.g., sensor systems, to prevent the identification of individuals. Rather than raw camera data being collected and processed by AV companies, the data could be processed on-vehicle via semantic segmentation or using obfuscation techniques.

3.3.4 Usable Privacy Solutions. Moreover, effective and strict privacy regulations on data collection and storage are needed as AV systems mature and become deployed in our society. Without these protections, end-users could resort to using anti-surveillance technologies/wearables that attack object detectors or sensors, increasing safety risks [136, 140].

> RESEARCH QUESTIONS—❶: *To what extent do AVs exacerbate surveillance? What are the direct implications of widespread AV adoption, and how will it impact personal freedoms? Can we provide privacy to the environment surrounding an AV without detriment to the AV's performance (safety and otherwise)?*

4 Sensor System

The perception modules in AVs are responsible for understanding and interpreting the environment. Typical mechanisms of self-driving vehicles involve perceiving surroundings, tracking objects, and planning future trajectories for the vehicle itself. As a result, the perception module is especially crucial as it provides initial results for use in safety-critical algorithms, which causes the perception module to be a desirable target for an attacker. On one hand, the adversary may alter the environment. The adversary may also physically inject a signal directly into the sensor. We categorize these as *Indirect* and *Direct AV Perception Attacks* respectively.

4.1 Indirect AV Perception Attacks

For *indirect* perception attacks, there are several proven methods from prior work. For instance, an adversary may introduce modifications to the road pavement to cause the vehicle to misinterpret its lane position and in turn mislead the lane detection system [70, 114]. Another strategy employed is to project adversarial patches to road signs and nearby walls [85, 88, 112]. The creation of adversarial objects has also shown effectiveness in misguiding object detection

systems [20]. Additionally, pedestrians may wear clothes with materials or patterns specifically designed to trick the sensors in order to remain undetected [148] or detected with incorrect label [14, 64, 140].

4.2 Direct AV Perception Attacks

When an adversary performs a *direct* perception attack, it involves physically manipulating the sensor to provide falsified measurements. These attacks may be naive with the only goal being to obfuscate the environment. However, an advanced adversary is capable of removing and/or adding specific objects without impacting the rest of the environment. These attacks have shown to at least reduce the overall performance of ML models that AVs use, or in the worst case cause a collision of the vehicle. One way of directly causing these attacks is for an adversary to use a laser to inject a desired signal into the sensor, such as LiDAR removal [18, 20, 56, 127, 150] and appearing [19, 21, 118, 124] attacks. To a lesser degree, lasers have been shown to be useful on other sensors as well, such as microphones [123]. While lasers have not been shown to be effective for these purposes on cameras [52], infrared light [112, 113, 130], acoustic signals [68], and electromagnetic interference [75] have been shown to be capable of directly injecting adversarial signals into the camera. Ultrasonic sensor spoofing has also been demonstrated to be feasible [141].

4.3 Defenses

To counter *indirect* perception attacks, state-of-the-art defenses suggest training the lane detection system on a diverse dataset that includes various perturbations, enabling the defensive module to recognize and ignore deceptive patterns. Adversarial training, a technique that has gained traction in recent years, specifically addresses attacks that use adversarial patches [26, 46, 85]. By training the vehicle's neural networks with adversarial examples, the system becomes more robust to deceptive inputs [124, 140, 149]. While this approach might marginally decrease the system's performance under normal conditions, its strength against known attacks is significantly enhanced.

Defenses against *direct* perception attacks are trickier due to a stronger and more resourceful adversary. These defenses may go as far as requiring hardware modifications to thwart the attacker. For instance, defenses for LiDAR removal attacks search for "shadows" left behind as an artifact of the attack [57, 130]. An adaptive adversary may attempt to hide the removal of an object by covering the shadow with a spoofed signal. There are also defenses for the appearing attacks, which take advantages of constraints imposed on the adversary for launching the attack [24, 137].

Filters are also used to defend against the *direct* attacks, such as a software filter for detecting infrared light attacks [130]. Hardware filters or shields are suggested for defense against attacks on MEMS accelerometers [126] and gyroscopes [51]. Physical filters may be useful to reduce the physical surface area an attack must target. For instance, camera hoods or diffracting films can block straight light beams from laser attacks [123].

Finally, a common approach is to use traditional intrusion detection systems to detect anomalous sensor data. Such a system may use estimates of sensor data to compare with the measurements [32, 33] or use a deep neural network to learn hidden features of anomalous sensor data [145].

4.4 Limitations & Takeaway

4.4.1 Absence of Provable Defenses. Often defenses against attacks on the perception system are constantly in a back-and-forth game of cat-and-mouse. Each time a new attack is defended against, a new attack takes its place to circumvent the defense. While defenses often remain under-evaluated, presented as an afterthought to the attack component, the primary issue is that defenses are typically hard to prove as secure. Oftentimes, defenses use probabilistic approaches that are still vulnerable to clever deception.

4.4.2 No Common Evaluation Practice. Despite the growing popularity of sensor attacks, the metrics used to evaluate success vary substantially across works. In some cases, simply achieving a misclassification is enough, while in other cases evading detection from the most popular defense is sufficient. While the difference in goal is largely due to the "story" a work builds, this makes comparing the impact of novel attacks impossible. This leads to our next takeaway.

4.4.3 Rarely Known Downstream Effect on Safety. The lack of common evaluation practice has caused a departure from focusing on actual AV safety as a consequence of a novel attack. While the attacks thoroughly evaluate the initial goal, it is not understood to what effect the attacks cause the AV to actually do anything dangerous. In a few work, dangerous control is shown to be feasible, but there is a notable absence of the range of scenarios for which this dangerous control may occur. Future work can improve the state of the art by providing a rigorous analysis of the state-space for which sensor attacks may successfully cause dangerous control outcomes.

> RESEARCH QUESTION—❷: *What set of evaluation metrics may satisfy the criteria for demonstrating novel AV perception attacks while simultaneously showing the real impact on the overall AV safety? Likewise, is it possible to find a provably secure defense against sensor attacks that do cause dangerous control outcomes?*

5 Compute Environment

The compute environment plays a vital role in AVs to guarantee the real-time completion and accuracy of the perception, planning, and control tasks [83]. Generally, it maintains the integrity of the communication middleware, real-time scheduling, and CPU/GPU resources.

5.1 AV Software Stack

Autoware is an "end-to-end" open-source software solution for autonomous driving [11]. It offers a comprehensive suite of tools and libraries for perception, planning, and control [10, 71]. A similar tool is Baidu Apollo, which is also open-sourced and aiming at level 4/5 autonomy [13]. However, incorrect implementations and misuse of APIs, concurrency, and memory expose many software vulnerabilities [45], which can leak privacy information and impact the AV's safety. One example is a cache side-channel attack can be used to infer the location of an AV that runs the adaptive Monte-Carlo localization (AMCL) algorithm [86].

The AVGuardian system is a testament to the importance of software defenses. It addresses overprivilege issues in AV software, a scenario where a software module has more permissions than it

requires, making it a potential weak point for attackers. By identifying and rectifying such instances, AVGuardian enforces stricter permission controls, thereby narrowing potential attack avenues [60]. Firmware code injection is another type of attack for the compute environment which could impact the perception and control modules of the AV [53]. By modifying the sensor data in real-time, the attacker achieves emergency braking or collision outcomes when running LiDAR-based perception and camera-LiDAR fusion perception in Apollo 7.0 [53]. Besides, by simply buffering and replaying scene data, safety-critical accidents are achieved.

5.2 Communication Middleware

While there exist several options for AV communication middleware, ROS is by far the most commonly used platform for autonomous vehicles [87, 105]. It faces significant security challenges, especially in its node and message communication systems. Within the ROS framework, there is no authentication mechanism for message passing or new node creation, which leaves it vulnerable to cyber-attacks. Attackers can use IP addresses and ports from the master node to create a new ROS node or hijack an existing one without additional authentication [67]. This vulnerability can lead to excessive consumption of system resources, such as memory and CPU usage, thereby negatively impacting the performance of legitimate ROS nodes and potentially causing the entire autonomous driving system to collapse. Additionally, ROS message security is compromised due to the ease of intercepting and replaying messages. Attackers can monitor and record ROS message traffic through the master node's IP address and port, storing this data in a ROS bag file. This allows them to replay historical messages, disrupting ongoing communications [67]. Furthermore, since ROS communication is based on socket communication, attackers can remotely monitor and intercept ROS messages without directly compromising the master node [79, 89]. ROS2 is developed using the Data Distribution Service (DDS) for security support [38], called Secure ROS2 (SROS2). However, the adversary can break the access control policies through out-of-date permission files or services. The reason is that when the access control policies are updated, SROS2 simply replaces the old permission files with the new ones, or sets up a new directory to store the new files and changes the corresponding load pointers [38].

5.3 Limitations & Takeaway

5.3.1 Lacking Comprehensive Code Check. The AV software stack is complex, encompassing a vast array of functionalities from sensor fusion and machine learning algorithms to decision-making and control systems. The large volume and intricacy of the code, coupled with the dynamic and interdependent nature of the systems, make it extremely difficult to thoroughly identify and rectify all potential bugs and errors. As a result, it leaves room for vulnerabilities that could be exploited by malicious attacks, thereby posing significant risks to the safety and reliability of AVs [45]. This scenario underscores a critical gap in the development and maintenance of AV software, highlighting the need for more robust and sophisticated methods in code analysis and security assurance. Formal verification could help [40]. Extensive testing and verification is needed for the algorithm and software stack.

5.3.2 Other Potential Vulnerabilities System Perspective. AVs confront an array of potential vulnerabilities from a system perspective, particularly concerning timing correctness and energy consumption, which are often ignored by the emphasis on accuracy and performance. Timing correctness is crucial in AVs as even slight deviations in processing times can lead to failures in decision-making, impacting the vehicle's ability to respond to real-time road situations [25]. Concurrently, energy consumption poses another challenge; while efforts to minimize power usage are essential for the longevity and efficiency of AVs, they must not compromise the system's operational integrity. Techniques like Dynamic Voltage and Frequency Scaling (DVFS), used for reducing energy consumption, can, if not carefully managed, impact the predictability and consistency of system performance.

> RESEARCH QUESTION—❸: *The complexity of AV software can lead to overlooked vulnerabilities, posing safety risks. Additionally, the emphasis on accuracy often neglects timing and energy efficiency aspects, which are critical for long-term safe AV performance. Can attacks on other attributes such as timing and energy lead to safety-critical failures?*

6 Algorithm

On top of the compute environment, algorithms are working with various sensing data to find the optimal navigation decision. To achieve this, the car should first track the trajectory of surrounding vehicles and predict/plan the future trajectory of the ego vehicle. Given that an erroneous navigation decision can directly lead to a fatal accident, we need to ensure safety from possible attacks or malfunctions arising from the driving environment.

As deep learning technology advances, recent AV algorithms have evolved to use Deep Neural Networks (DNNs). Accordingly, there is significant research at the algorithm level that demonstrates a strong adversary exploiting the probabilistic nature of DNNs to intentionally cause safety-critical failures of AV decision-making processes [142].

6.1 Tracking

3D Multi-object tracking (MOT) is achieved by capturing object motion across detection results in consecutive frames [135]. Due to its ability to interpret historical context, tracking is increasingly used as a defensive strategy against attacks targeting AV perception systems. For example, PercepGuard [88] checks the consistency of tracked objects with their respective category characteristics, acting as a countermeasure against misclassification attacks. In a similar vein, 3D-TC2 [143] utilizes tracking and prediction of objects' future positions to check temporal consistency with detected objects, effectively countering LiDAR spoofing attacks.

Concurrently, advanced attacks specifically designed for tracking algorithms have emerged. Jia *et al.* [69] introduced an adversarial attack and defense algorithm that manipulates the location and size of bounding boxes to boost digital domain robustness. Likewise, Muller *et al.* [93] have devised an innovative tracker hijacking method. This technique modifies the trajectory of a target object by either shifting it onto an unrelated path or completely altering its intended

trajectory. A key feature of this method is the creation of physically plausible attacks, avoiding unrealistic manipulations, such as moving objects into the sky.

Moreover, tracking data is highly susceptible to malfunctions due to challenges like occlusion, dynamic and unpredictable object movements, and variations in the ego vehicle's perspective. The absence of depth information in vision data often leads to misinterpretations. As a result, continuous research into attacks, defenses, and potential malfunctions is essential for ensuring the safety and security of AV systems.

6.2 Navigational Safety

In the planning module, AV systems are centered around finding the most efficient routes to their destinations. This involves merging data from detection and tracking systems to formulate future trajectories and actions. In this realm, Reinforcement Learning (RL) is notably effective, optimizing behavior across varied driving conditions while focusing on maximizing predefined rewards. Consequently, RL-based planning algorithms have increasingly become targets for adversaries. For instance, He *et al.* [59] strategically manipulate observed states and environmental dynamics to mislead RL-based decision-making systems.

Taking a different direction, ACERO [122] unveils a novel attack strategy wherein an adversary orchestrates their vehicle movements to endanger a targeted vehicle. This technique involves crafting an optimal adversarial driving maneuvers by considering safety requirements and physical constraints for a successful attack execution.

Considering that navigation decisions are more closely tied to a vehicle's active decision-making than any other module, a heightened level of safety awareness is essential. Mo *et al.* [92], for instance, showcase the integration of risk assessment and safe policy search in RL agents. In a similar context, Li *et al.* [82] contribute to a more realistic risk assessment in multi-AV environments by taking into account the various driving intentions of surrounding vehicles, such as cut-ins and overtakings.

6.3 Robustness to Domain Shifts

One critical problem in self-driving vehicles is ensuring consistent object detection or segmentation performance across different domain shifts such as weather. Ensuring the safety of self-driving vehicles against adverse weather conditions such as rain, fog, or snow is a difficult yet important problem to avoid unexpected behavior in such conditions. The importance of this problem has seen an emergence of datasets that include different weather conditions such as BDD100K [144], Cityscapes [27] and its rainy [63] and foggy [110] variants, KITTI [47] and its rainy [91] and foggy [91] variants, and NuScenes [16]. With such datasets, some AV specific approaches have been proposed to handle different weather conditions [8, 81, 129]. However, it is worth looking at advances in the more general computer vision domain to understand how such approaches may be improved.

As a case study, unsupervised domain adaptation in object detection and segmentation [23, 44, 61, 62, 73, 81, 96, 99, 100, 107, 109, 119, 147] is one key area of interest due to its ability to leverage unlabeled images or synthetically generated images. These

approaches include adversarial learning to learn domain invariant features [23, 44, 61, 100, 109, 119], self-training on pseudo labels [73, 107, 147], image-to-image translation [8, 62], and learning with motion or temporal-based priors [50, 80, 101, 139], to name a few. Many of these works use simulated or real AV datasets and focus on the machine learning aspects with evaluations aimed around mIoU for segmentation and mAP for detection. These approaches from the vision domain offer much potential inspiration towards building AV perception systems that can one day be trusted to handle the whole range of realistic driving environments.

6.4 Limitations & Takeaway

6.4.1 Error Propagation in Sequential Execution of Multi-DNN. The perception algorithm plays a pivotal role as it provides initial results to subsequent modules. Given that state-of-the-art perception is not entirely accurate in its predictions, there is a risk of error propagation to further "downstream" modules. To address this, some algorithms incorporate strategies to validate perception results. These strategies involve statistically calculating the propagated error of bounding boxes [133] or running algorithms at the raw sensor data level [24]. However, accurately computing data uncertainty, intra-model uncertainty, and inter-model uncertainty remains a complex challenge. This complexity necessitates a more comprehensive understanding to ensure a robust risk assessment.

6.4.2 Limited Coverage on Complex and Diverse Real-World Driving Environment. Self-driving companies train foundational models for different tasks and deploy these models in cars via post-optimization [115, 132]. A significant challenge arises from the limited datasets used in training these models, which may not encompass the full spectrum of real-world road environments. This limitation often results in a distribution shift in the inputs fed into AV systems, leading to erroneous outputs.

To mitigate these shift issues, methods such as fine-tuning, domain generalization, and test-time adaptation are being explored [8, 81, 129]. These methods aim to lessen the impact of domain shifts and ensure that systems remain up-to-date. However, since machine learning models must be optimized before deployment, this optimization process inevitably limits model size, consequently affecting their learning capabilities. Therefore, a crucial decision lies in balancing the trade-off between optimization and learning capability, and determining which model (deployed vs. foundational) should be adjusted, and at what point in time.

6.4.3 Lack of AV Specific Solutions for Domain Shift Invariance. To guarantee the safety of AVs, the perception algorithm must be robust to distributional shifts such as poor weather or other environmental conditions that a vehicle may run into [8, 23, 61, 62, 73, 81, 99, 100, 107, 119, 129, 147]. The robustness against different distribution-shifts is a defense against a "natural" attacker, where the threat vector is to identify the worst realistic driving conditions for the perception algorithm. Such work includes papers that are marketed more overtly towards AVs [8, 81, 129] and others that are more general, domain-adaptation focused but evaluated on AV-related datasets [23, 50, 61, 62, 73, 80, 100, 101, 107, 119, 139, 147]. We highlight two interesting trends that could be useful to consider for further research.

The first is that these approaches generally focus on the RGB image domain and use general techniques that do not make strong assumptions about the AV's setting. For example, incorporating general state-of-the-art vision techniques such as domain adaptation with multi-sensor fusion networks (assuming of the presence of LiDAR) may lead to more robust feature extraction specifically for AVs. Traffic signs also have less variation than ImageNet [108] classes (e.g., traffic signs are limited to rigid shapes and are usually never upside-down whereas a dog has a lot more variation in its shape). This may lead to AV specific solutions to extracting robust, domain invariant features, for example by extending approaches in hierarchical classification where the root classifier determines the shape and the leaf classifiers determine the specific class [22]. Such an approach could bias a model towards features that are more invariant to weather conditions and may work better on traffic sign classification than more general datasets like ImageNet [108].

Second, these works evaluate mostly on the basis of mAP or mIoU, which are more vision-aligned metrics and do not necessarily capture the safety risk of the overall system that may be using these ML models. The real-world ramifications of the robustness (or lack thereof) may require AV-specific metrics that take into account the full system and estimates the end-to-end safety implications of any misaligned predictions. For example, incorrectly detected objects could be weighted by their distance to the AV, with closer objects being penalized harsher since farther objects can still be avoided if detected properly in a near-future frame. Specific misclassifications can also be less severe than others (e.g., mistaking a stop sign for a speed limit sign vs. mistaking a cat for a dog). Finally, the performance of a perception algorithm can be measured in the context of how it performs within end-to-end AV system (i.e., planning and control).

> RESEARCH QUESTION–❹: *How can we reinforce robustness and reliability, both intra-module and inter-module, within our pipeline to ensure preparedness for increasingly complex and unpredictable driving scenarios? Moreover, can such preparation increase defenses against a strong adversary?*

7 Environment Interaction

7.1 Legal Considerations

In the context of an AV attack impacting human safety, the legal implications are multifaceted and complex. When such an attack results in a collision or similar incident, determining liability involves analyzing several factors, including the nature of the ML system, the intent behind the adversarial input, and existing legal frameworks.

First, the nature of the ML system plays a crucial role. For AVs, the manufacturer could be held liable under product liability laws if the system is found to be defective or not reasonably safe. Vladeck argues that traditional product liability principles can extend to AI systems, particularly when they fail to perform as safely as an ordinary consumer would expect [128]. Similarly, Calo suggests that manufacturers might be held strictly liable for damages caused by their products, regardless of fault or negligence [17].

Second, the source and intent of the adversarial input are crucial. If the AV attack is traced back to a specific entity or individual, they could be held liable under tort law for intentionally causing

harm. As Kerr points out, if a person knowingly manipulates data to cause an AI system to malfunction, they might be liable for any resulting damages or injuries under theories of intentional torts or cyber-trespass [72].

Finally, the existing legal frameworks and their adaptation to emerging technologies are paramount in such scenarios. As Pagallo discusses, the evolution of AI challenges traditional legal concepts of liability and responsibility, necessitating a reevaluation of legal frameworks to accommodate these technological advances [98]. This view is also supported by Solaiman, who highlights the need for new legal paradigms to address the unique challenges posed by ML systems [121], and Widen and Koopman, who propose that the AV industry should cooperate with government regulators to slowly introduce policy that would improve public comfort with AVs [134].

7.2 Regulatory Standards

There are several standards that govern the safety and security of autonomous vehicles. The most notable standard is ANSI/UL 4600, which outlines a standard for autonomous product safety [5]. While it does not define pass/fail criteria for safety, it does provide guidance for how to ensure safety is applied to the design, testing, validation, human-machine interaction, and other processes. Alongside this standard, the NHTSA requires reporting of crashes that involve vehicles that are equipped with autonomous systems [95]. Moreover, there are standards from ISO that define the functional safety of vehicles in the event of system failures [3], as well as the intended functionality of vehicle systems in the absence of faults [2]. The IEEE also provides a standard set of scenarios that AVs should be able to safely handle [120].

For security, the primary standard for security of road vehicles is ISO/SAE 21434 [1]. There are other standards that concern the standard for vehicle cybersecurity management systems [106], and auditing the cybersecurity of vehicles [6]. Outside of vehicles, there are additional standards for security, such as ISO27000 [4] and ANSI/ISA 62443 [7].

7.3 Limitations & Takeaway

7.3.1 Laws Must Balance Safety Without Slowing Tech Growth. Due to recent safety concerns regarding AVs, there is a growing distrust for their public use. In response to these concerns, governments (local or national) may begin to impose strict sanctions on their operation. While this response is well-intentioned and likely required in order to regain faith in the public deployment of AVs, there is room for the legislation that follows to strongly discourage the development of the tech. If it is too strict, then the barriers it enforces may deincentivize further investment and ultimately kill the technology on the stop. Future legislation must carefully consider the intended outcomes once enacted, and support should be provided to AV companies to ensure they can continue development.

7.3.2 Ad Hoc Provisions of AV "Driving Licenses". The existing standards primarily serve the purpose of L2/L3 AV features. Presently, when a L4 AV hits the road, local governments provide a license to the autonomous driver based on *ad hoc* standards that adapt as issues arise. This method of providing fixes as issues arise is against the goals of safety engineering, which seeks to eliminate risk before

damages occur. Presently, issuing licenses to L4 AVs require stricter regulation and a stronger safety culture.

7.3.3 Modeling Human Behavior is Hard. Despite the legal necessity to demonstrate functional safety, especially when human-life is at risk, it is extremely difficult to model the behavior of other people. This includes how other vehicles may respond to the ego vehicle's control, how aggressive cyclists may behave on the road, or how pedestrians may act during a road crossing scenario. No matter the situation, people can sometimes do seemingly random things and catch the AV off-guard. As a result, despite any best effort, it is not currently practical to completely guarantee functional safety requirements.

> RESEARCH QUESTION—❺: *Future legislation will need to consider how and when AV software can be granted a license to autonomously operate in specific environments. How can this be provided to ensure the public does not involuntarily partake in a dangerous experiment, while simultaneously encourage the further development of AV technology?*

8 End-to-End Evaluation Examples

Before moving on to our future recommendations, we discuss some cases where there is a substantial attempt at end-to-end testing of AVs. One of the earliest models of AVs is from Mobileye's RSS formal model for providing safety assurances [117]. RSS provides a model based on physics for assuring the safety of the vehicle. While the model is abstract and requires exact precision of the measurements of the vehicle's surroundings to ensure the safety of the vehicle, its a step in the right direction. Other work investigate impacts on safety due to failures in the perception stack to fill this gap [43]. There are also additions to end-to-end testing that attempt to test for rare events [97], using adaptive stress testing to pursue stronger coverage of scenarios [28], and using fuzzing to discover bugs in the AV stack [74].

The major gap in end-to-end validation testing of AV safety is in the simulation tools. There is a substantial gap between AVs in simulation compared to real-world AVs, otherwise known as the sim-to-real gap. Overcoming this drawback is still an area of active inquiry. Current solutions primarily seek to cross the sim-to-real gap via transferability of training in simulation to real-world environments [36, 111, 125]. These usages of sim-to-real do not seek to validate the system safety.

9 Future Research Recommendations

Recommendation #1: For future research in AV security, a key recommendation is to focus on comprehensive end-to-end testing strategies that encompass the entire system's operation, from perception to control, to ensure robustness and safety in diverse real-world scenarios. This would include the future creation of universal metrics that may be used across all problems to accurately weigh the strengths and weaknesses of various approaches. These metrics may measure various aspects of the pipeline in order to demonstrate where attacks/defenses are most effective, as well as the end result so that one may see the threat that an attack has on functional safety.

Recommendation #2: Effectively utilizing the data from surrounding vehicles and integrating it is essential to gain a more comprehensive understanding of the driving environment. Cooperative perception and planning strategies can significantly enhance safety and security, particularly in scenarios where a vehicle is compromised or its sensors are disrupted. This approach ensures more reliable and accurate sensing outcomes, even in challenging circumstances. Moreover, it's crucial to consider that drivers themselves can inadvertently be a source of safety and security issues. Even without malicious intent, their unavoidable actions may resemble attack-like behavior to other vehicles in specific traffic situations. Therefore, in addition to the AV technology stacks illustrated in Fig. 1, monitoring and understanding driver behaviors can play a pivotal role in preventing AVs from exacerbating potential hazards.

Recommendation #3: The licenses of AVs should have more concrete and strict regulation for when and how they are provided and enforced. This would include clearer standards for the sort of scenarios that the AVs can provide provable functional safety for. Moreover, these AVs should be required to have well-regulated privacy policies with respect to the data they collect on the road. Packaging privacy requirements with the safety requirements of an AV license would provide an encouraging solution to both issues at once. What remains to be determined is where the regulation "sweet spot" is for strictly enforcing an AV license. Government should seek to avoid deincentivizing the development of AVs, so finding this balance is crucial.

References

[1] ISO 21434. 2021. *Road vehicles - Cybersecurity engineering.* Standard. International Organization for Standardization.

[2] ISO 21448. 2022. *Road vehicles – Safety of the intended functionality.* Standard. International Organization for Standardization.

[3] ISO 26262. 2018. *Road vehicles – Functional safety.* Standard. International Organization for Standardization.

[4] ISO 27000. 2018. *Security techniques - Information security management systems.* Standard. International Organization for Standardization.

[5] ANSI/UL 4600. 2023. *Standard for Safety for the Evaluation of Autonomous Products.* Standard. American National Standards Institute.

[6] ISO/PAS 5112. 2022. *Road vehicles - Guidelines for auditing cybersecurity engineering.* Standard. International Organization for Standardization.

[7] ISA/IEC 62443. 2020. *Security of Industrial Automation and Control Systems.* Standard. American National Standards Institute.

[8] Sabbir Ahmed, Uday Kamal, and Md Kamrul Hasan. 2021. DFR-TSD: A deep learning based framework for robust traffic sign detection under challenging weather conditions. *IEEE Transactions on Intelligent Transportation Systems* 23, 6 (2021), 5150–5162.

[9] ArduPilot. 2023. ArduPilot. https://ardupilot.org/ardupilot/index.html. [Online; accessed 23-Dec-2023].

[10] Autoware. 2023. Autoware High-level architecture design. https://autowarefoundation.github.io/autoware-documentation/main/design/autoware-architecture/. [Online; accessed 23-July-2023].

[11] Why Autoware. 2023. Autoware High-level architecture design. https://autoware.org/about/why-autoware/. [Online; accessed 26-July-2023].

[12] Baidu. 2023. Apollo Cyber RT. https://cyber-rt.readthedocs.io/en/latest/. [Online; accessed 23-Dec-2023].

[13] Baidu. 2023. Apollo Open Platform. https://developer.apollo.auto/. [Online; accessed 17-Dec-2023].

[14] Pesala Bandara. 2023. *This Clothing Line Tricks AI Cameras Without Covering Your Face.* PetaPixel. https://web.archive.org/web/20231206174711/https://petapixel.com/2023/01/20/this-clothing-line-tricks-ai-cameras-without-covering-your-face/

[15] Cara Bloom, Joshua Tan, Javed Ramjohn, and Lujo Bauer. 2017. Self-driving cars and data collection: Privacy perceptions of networked autonomous vehicles. In *Proceedings of the Thirteenth Symposium on Usable Privacy and Security (SOUPS '17).* USENIX, Berkeley, CA, USA, 357–375.

[16] Holger Caesar, Varun Bankiti, Alex H. Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. 2020. nuScenes: A Multimodal Dataset for Autonomous Driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).* IEEE, New York, NY, USA, 11621–11631.

[17] Ryan Calo. 2015. Robotics and the Lessons of Cyberlaw. *California Law Review* 103, 3 (2015), 513–563.

[18] Yulong Cao, S. Hrushikesh Bhupathiraju, Pirouz Naghavi, Takeshi Sugawara, Z. Morley Mao, and Sara Rampazzi. 2023. You Can't See Me: Physical Removal Attacks on LiDAR-based Autonomous Vehicles Driving Frameworks. In *The Proceedings of the 32nd USENIX Security Symposium (USENIX Security '23)*. USENIX, Berkeley, CA, USA, 2993–3010.

[19] Yulong Cao, Jiaxiang Ma, Kevin Fu, Rampazzi Sara, and Morley Mao. 2021. Demo: Automated Tracking System For LiDAR Spoofing Attacks On Moving Targets. In *Workshop on Automotive and Autonomous Vehicle Security 2021 (AutoSec '21)*. ISOC, Reston, VA, USA, 1 pages.

[20] Yulong Cao, Ningfei Wang, Chaowei Xiao, Dawei Yang, Jin Fang, Ruigang Yang, Qi Alfred Chen, Mingyan Liu, and Bo Li. 2021. Invisible for both Camera and LiDAR: Security of Multi-Sensor Fusion based Perception in Autonomous Driving Under Physical-World Attacks. In *Proceedings of the 42nd IEEE Symposium on Security and Privacy (S&P '21)*. IEEE, New York, NY, USA, 176–194.

[21] Yulong Cao, Chaowei Wiao, Benjamin Cyr, Yimeng Zhou, Won Park, Sara Rampazzi, Qi Alfred Chen, Kevin Fu, and Z. Morley Mao. 2019. Adversarial Sensor Attack on LiDAR-based Perception in Autonomous Driving. In *The Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security (CCS '19)*. ACM, New York, NY, USA, 2267–2281.

[22] Varun Chandrasekaran, Brian Tang, Nicolas Papernot, Kassem Fawaz, Somesh Jha, and Xi Wu. 2019. *Rearchitecting Classification Frameworks For Increased Robustness.* University of Wisconsin-Madison. arXiv:1905.10900 [cs.LG]

[23] Yuhua Chen, Wen Li, Christos Sakaridis, Dengxin Dai, and Luc Van Gool. 2018. Domain Adaptive Faster R-CNN for Object Detection in the Wild. In *The Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, New York, NY, USA, 3339–3348.

[24] Minkyoung Cho, Yulong Cao, Zixiang Zhou, and Z. Morley Mao. 2023. ADoPT: LiDAR Spoofing Attack Detection Based on Point-Level Temporal Consistency. In *The Proceedings of the 34th British Machine Vision Conference (BMVC '23)*. BMVA, Durham, U.K., 13 pages.

[25] Minkyoung Cho and Kang G. Shin. 2023. *DynaMIX: Resource Optimization for DNN-Based Real-Time Applications on a Multi-Tasking System.* University of Michigan. arXiv:2302.01568 [cs.LG]

[26] Jeremy Cohen, Elan Rosenfeld, and Zico Kolter. 2019. Certified Adversarial Robustness via Randomized Smoothing. In *The Proceedings of the 36th International Conference on Machine Learning (ICML '19)*. PMLR, New York, NY, USA, 1310–1320.

[27] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. 2015. The Cityscapes Dataset for Semantic Urban Scene Understanding. In *The Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, New York, NY, USA, 3213–3223.

[28] Anthony Corso, Peter Du, Katherine Driggs-Campbell, and Mykel J. Kochenderfer. 2019. Adaptive Stress Testing with Reward Augmentation for Autonomous Vehicle Validation. In *The Proceedings of the IEEE 22nd International Conference on Intelligent Transportation Systems (ITSC '19)*. IEEE, New York, NY, USA, 163–168.

[29] Cruise. 2022. *Cruise Safety Report.* Cruise. https://assets.ctfassets.net/95kuvdv8zn1v/zKJHD7X22fNzpAJztpd5K/ac6cd2419f2665000e4eac3b7d16ad1c/Cruise_Safety_Report_2022_sm-optimized.pdf

[30] Cruise. 2023. *Cruise Safety.* Cruise. https://getcruise.com/safety/

[31] Cruise. 2023. *Important Updates from Cruise.* Cruise. https://getcruise.com/news/blog/2023/important-updates-from-cruise/

[32] Noah T. Curran, Arun Ganesan, Mert D. Pesé, and Kang G. Shin. 2023. Using Phone Sensors to Augment Vehicle Reliability. In *The Proceedings of the 2023 IEEE Conference on Communications and Network Security (CNS '23)*. IEEE, New York, NY, USA, 9 pages.

[33] Noah T. Curran, Thomas W. Kennings, and Kang G. Shin. 2024. Analysis and Prevention of MCAS-Induced Crashes. In *The Proceedings of the 2024 ACM SIGBED International Conference on Embedded Software (EMSOFT '24)*. IEEE, New York, NY, USA, 12 pages.

[34] Noah T. Curran, Kang G. Shin, William Hass, Lars Wolleschensky, Rekha Singoria, Isaac Snellgrove, and Ran Tao. 2023. WIP: Augmenting Vehicle Safety With Passive BLE. In *The Proceedings of the Inaugural Symposium on Vehicle Security and Privacy (VehicleSec '23)*. ISOC, Reston, VA, USA, 5 pages.

[35] Laurent David, Francis Cottet, and Nimal Nissanke. 2001. Jitter Control in On-line Scheduling of Dependent Real-Time Tasks. In *The Proceedings 22nd IEEE Real-Time Systems Symposium (RTSS '01)*. IEEE, New York, NY, USA, 49–58.

[36] Iván García Daza, Rubén Izquierdo, Luis Miguel Martínez, Ola Benderius, and David Fernández Llorca. 2023. Sim-to-real transfer and reality gap modeling in model predictive control for autonomous driving. *Applied Intelligence* 53, 10 (2023), 12719–12735.

[37] Dionisio de Niz, Karthik Lakshmanan, and Ragunathan Rajkumar. 2009. On the Scheduling of Mixed-Criticality Real-Time Task Sets. In *The Proceedings of the 2009 30th IEEE Real-Time Systems Symposium (RTSS '09)*. IEEE, New York, NY, USA, 291–300.

[38] Gelei Deng, Guowen Xu, Yuan Zhou, Tianwei Zhang, and Yang Liu. 2022. On the (In)Security of Secure ROS2. In *The Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security (CCS '22)*. ACM, New York, NY, USA, 739–753.

[39] Arvind Easwaran and Bjorn Andersson. 2009. Resource Sharing in Global Fixed-Priority Preemptive Multiprocessor Scheduling. In *The Proceedings of the 2009 30th IEEE Real-Time Systems Symposium (RTSS '09)*. IEEE, New York, NY, USA, 377–386.

[40] Kathleen Fisher, John Launchbury, and Raymond Richards. 2017. The HACMS program: using formal methods to eliminate exploitable bugs. *Philosophical Transactions of the Royal Society A* 375 (2017), 18 pages. Issue 2104.

[41] Foresight. 2023. An Overview of Autonomous Sensors – LIDAR, RADAR, and Cameras. https://www.foresightauto.com/an-overview-of-autonomous-sensors-lidar-radar-and-cameras/. [Online; accessed 27-Dec-2023].

[42] Mozilla Foundation. 2023. *Privacy Nightmare on Wheels: Every Car Brand Reviewed By Mozilla - Including Ford, Volkswagen and Toyota - Flunks Privacy Test.* Mozilla Foundation. https://foundation.mozilla.org/en/blog/privacy-nightmare-on-wheels-every-car-brand-reviewed-by-mozilla-including-ford-volkswagen-and-toyota-flunks-privacy-test

[43] Daniel J. Fremont, Edward Kim, Yash Vardhan Pant, Sanjit A. Seshia, Atul Acharya, Xantha Bruso, Paul Wells, Steve Lemke, Qiang Lu, and Shalin Mehta. 2020. Formal Scenario-Based Testing of Autonomous Vehicles: From Simulation to the Real World. In *The Proceedings of the 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC '20)*. IEEE, New York, NY, USA, 8 pages.

[44] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario March, and Victor Lempitsky. 2016. Domain-Adversarial Training of Neural Networks. *Journal of Machine Learning Research* 17, 59 (2016), 1–35.

[45] Joshua Garcia, Yang Feng, Junjie Shen, Sumaya Almanee, Yuan Xia, and Qi Alfred Chen. 2020. A Comprehensive Study of Autonomous Vehicle Bugs. In *The Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering (ICSE '20)*. ACM, New York, NY, USA, 385–396.

[46] Timon Gehr, Matthew Mirman, Dana Drachsler-Cohen, Petar Tsankov, Swarat Chaudhuri, and Martin Vechev. 2018. AI2: Safety and Robustness Certification of Neural Networks with Abstract Interpretation. In *2018 IEEE Symposium on Security and Privacy (S&P '18)*. IEEE, New York, NY, USA, 3–18.

[47] Andreas Geiger, Philip Lenz, and Raquel Urtasun. 2012. Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite. In *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR '12)*. IEEE, New York, NY, USA, 3354–3361.

[48] Soheila Ghane, Alireza Jolfaei, Lars Kulik, Kotagiri Ramamohanarao, and Deepak Puthal. 2020. Preserving Privacy in the Internet of Connected Vehicles. *IEEE Transactions on Intelligent Transportation Systems* 22, 8 (2020), 5018–5027.

[49] Dorothy J. Glancy. 2012. Privacy in Autonomous Vehicles. *Santa Clara Law Review* 52 (2012), 1171–1239.

[50] Dayan Guan, Jiaxing Huang, Aoran Xiao, and Shijian Lu. 2021. Domain Adaptive Video Segmentation via Temporal Consistency Regularization. In *The Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, New York, NY, USA, 8053–8064.

[51] Mordechai Guri. 2021. GAIROSCOPE: Leaking Data from Air-Gapped Computers to Nearby Smartphones using Speakers-to-Gyro Communication. In *2021 18th International Conference on Privacy, Security and Trust (PST '21)*. IEEE, New York, NY, USA, 10 pages.

[52] R. Spencer Hallyburton, Yupei Liu, Yulong Cao, Z. Morley Mao, and Miroslav Pajic. 2022. Security Analysis of Camera-LiDAR Fusion Against Black-Box Attacks on Autonomous Vehicles. In *The Proceedings of the 31st USENIX Security Symposium (USENIX Security '22)*. USENIX, Berkeley, CA, USA, 1903–1920.

[53] R. Spencer Hallyburton, Qingzhao Zhang, Z. Morley Mao, and Miroslav Pajic. 2023. *Partial-Information, Longitudinal Cyber Attacks on Sensing in Autonomy.* Duke University. arXiv:2303.03470 [cs.CR]

[54] Moncef Hamdaoui and Parameswaran Ramanathan. 1995. A Dynamic Priority Assignment Technique for Streams with (m,k)-Firm Deadlines. *IEEE Trans. Comput.* 44, 12 (1995), 1443–1451.

[55] Muneeb Ul Hassan, Mubashir Husain Rehmani, and Jinjun Chen. 2020. Differential Privacy Techniques for Cyber Physical Systems: A Survey. *IEEE Communications Surveys & Tutorials* 22, 1 (2020), 746–789.

[56] Zhongyuan Hau, Kenneth T. Co, Soteris Demetriou, and Emil C. Lupu. 2021. Object Removal Attacks on LiDAR-based 3D Object Detectors. In *Workshop on Automotive and Autonomous Vehicle Security 2021 (AutoSec '21)*. ISOC, Reston, VA, USA, 4 pages.

[57] Zhongyuan Hau, Soteris Demetriou, and Emil C. Lupu. 2022. Using 3D Shadows to Detect Object Hiding Attacks on Autonomous Vehicle Perception. In *2022 IEEE Security and Privacy Workshops (SPW '22)*. IEEE, New York, NY, USA, 229–235.

[58] Andrew J. Hawkins. 2023. *Robotaxis are driving on thin ice.* The Verge. https://www.theverge.com/2023/8/15/23831170/robotaxi-cpuc-sf-waymo-cruise-traffic-halt

[59] Xiangkun He, Wenhui Huang, and Chen Lv. 2024. Toward Trustworthy Decision-Making for Autonomous Vehicles: A Robust Reinforcement Learning Approach with Safety Guarantees. *Engineering* 33 (2024), 77–89.

[60] David Ke Hong, John Kloosterman, Yuqi Jin, Yulong Cao, Qi Alfred Chen, Scott Mahlke, and Z. Morley Mao. 2020. AVGuardian: Detecting and Mitigating Publish-Subscribe Overprivilege for Autonomous Vehicle Systems. In *2020 IEEE European Symposium on Security and Privacy (EuroS&P)*. IEEE, New York, NY, USA, 445–459.

[61] Cheng-Chun Hsu, Yi-Hsuan Tsai, Yen-Yu Lin, and Ming-Hsuan Yang. 2020. Every Pixel Matters: Center-aware Feature Alignment for Domain Adaptive Object Detector. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX 16*. Springer, New York, NY, USA, 733–748.

[62] Han-Kai Hsu, Chun-Han Yao, Yi-Hsuan Tsai, Wei-Chih Hung, Hung-Yu Tseng, Maneesh Singh, and Ming-Hsuan Yang. 2020. Progressive Domain Adaptation for Object Detection. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*. IEEE, New York, NY, USA, 749–757.

[63] Xiaowei Hu, Chi-Wing Fu, Lei Zhu, and Pheng-Ann Heng. 2019. Depth-Attentional Features for Single-Image Rain Removal. In *The Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, New York, NY, USA, 8022–8031.

[64] Zhanhao Hu, Siyuan Huang, Xiaopei Zhu, Fuchun Sun, Bo Zhang, and Xiaolin Hu. 2022. Adversarial Texture for Fooling Person Detectors in the Physical World. In *Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, New York, NY, USA, 13307–13316.

[65] Henry Alexander Ignatious, Hesham El-Sayed, and Manzoor Ahmed Khan. 2022. An overview of sensors in Autonomous Vehicles. *Procedia Computer Science* 198 (2022), 736–741.

[66] Deepthi Jallepalli, Navya Chennagiri Ravikumar, Poojitha Vurtur Badarinath, Shravya Uchil, and Mahima Agumbe Suresh. 2021. Federated Learning for Object Detection in Autonomous Vehicles. In *2021 IEEE Seventh International Conference on Big Data Computing Service and Applications (BigDataService)*. IEEE, New York, NY, USA, 107–114.

[67] Se-Yeon Jeong, I-Ju Choi, Yeong-Jin Kim, Yong-Min Shin, Jeong-Hun Han, Goo-Hong Jung, and Kyoung-Gon Kim. 2017. A Study on ROS Vulnerabilities and Countermeasure. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, New York, NY, USA, 147–148.

[68] Xiaoyu Ji, Yushi Cheng, Yuepeng Zhang, Kai Wang, Chen Yan, Wenyuan Xu, and Kevin Fu. 2021. Poltergeist: Acoustic Adversarial Machine Learning against Cameras and Computer Vision. In *2021 IEEE Symposium on Security and Privacy (SP)*. IEEE, New York, NY, USA, 160–175.

[69] Shuai Jia, Chao Ma, Yibing Song, and Xiaokang Yang. 2020. Robust Tracking against Adversarial Attacks. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIX 16*. Springer, New York, NY, USA, 69–84.

[70] Pengfei Jing, Qiyi Tang, Yuefeng Du, Lei Xue, Xiapu Luo, Ting Wang, Sen Nie, and Shi Wu. 2021. Too Good to Be Safe: Tricking Lane Detection in Autonomous Driving with Crafted Perturbations. In *30th USENIX Security Symposium (USENIX Security 21)*. USENIX, Berkeley, CA, USA, 3237–3254.

[71] Shinpei Kato, Shota Tokunaga, Yuya Maruyama, Seiya Maeda, Manato Hirabayashi, Yuki Kitsukawa, Abraham Monrroy, Tomohito Ando, Yusuke Fujii, and Takuya Azumi. 2018. Autoware on Board: Enabling Autonomous Vehicles with Embedded Systems. In *2018 ACM/IEEE 9th International Conference on Cyber-Physical Systems (ICCPS)*. IEEE, New York, NY, USA, 287–296.

[72] Orin S Kerr. 2016. Norms of Computer Trespass. In *116 Columbia Law Review 1143*, Vol. 116. HeinOnline, Buffalo, NY, USA, 42 pages.

[73] Mehran Khodabandeh, Arash Vahdat, Mani Ranjbar, and William G. Macready. 2019. A Robust Learning Approach to Domain Adaptive Object Detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, New York, NY, USA, 480–490.

[74] Seulbae Kim, Major Liu, Junghwan "John" Rhee, Yuseok Jeon, Yonghwi Kwon, and Chung Hwan Kim. 2022. DriveFuzz: Discovering Autonomous Driving Bugs through Driving Quality-Guided Fuzzing. In *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security (CCS '22)*. ACM, New York, NY, USA, 1753−−1767.

[75] Sebastian Köhler, Richard Baker, and Ivan Martinovic. 2022. Signal Injection Attacks against CCD Image Sensors. In *Proceedings of the 2022 ACM on Asia Conference on Computer and Communications Security (ASIACCS)*. ACM, New York, NY, USA, 294–308.

[76] Philip Koopman. 2023. *The Cruise Safety Stand-Down – What Happens Next?* Safe Autonomy. https://safeautonomy.blogspot.com/2023/10/the-cruise-safety-stand-down-and-what.html

[77] Philip Koopman. 2023. *Safety Analysis of Two Cruise Robotaxi Pedestrian Injuries*. Safe Autonomy. https://safeautonomy.blogspot.com/2023/10/safety-analysis-of-two-cruise-robotaxi.html

[78] Luis F Alvarez León. 2019. Eyes on the Road: Surveillance Logics in the Autonomous Vehicle Economy. *Surveillance & Society* 17, 1/2 (2019), 198–204.

[79] Francisco Javier Rodrıguez Lera, Jesús Balsa, Fernando Casado, Camino Fernández, Francisco Martın Rico, and Vicente Matellán. 2016. Cybersecurity in Autonomous Systems: Evaluating the performance of hardening ROS. *Málaga, Spain* 47 (2016), 7 pages.

[80] Hanhan Li, Ariel Gordon, Hang Zhao, Vincent Casser, and Anelia Angelova. 2021. Unsupervised Monocular Depth Learning in Dynamic Scenes. In *Conference*

[81] Jinlong Li, Runsheng Xu, Jin Ma, Qin Zou, Jiaqi Ma, and Hongkai Yu. 2023. Domain Adaptive Object Detection for Autonomous Driving under Foggy Weathe. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. IEEE, New York, NY, USA, 612–622.

[82] Lin Li, Wanzhong Zhao, and Chunyan Wang. 2022. POMDP motion planning algorithm based on multi-modal driving intention. *IEEE Transactions on Intelligent Vehicles* 8, 2 (2022), 1777–1786.

[83] Liangkai Liu, Sidi Lu, Ren Zhong, Baofu Wu, Yongtao Yao, Qingyang Zhang, and Weisong Shi. 2020. Computing Systems for Autonomous Driving: State-of-the-Art and Challenges. *IEEE Internet of Things Journal* 8, 8 (2020), 6469–6486.

[84] Julia Love. 2023. Self-Driving Car Footage Gives Police Controversial New Tool. https://news.bloomberglaw.com/white-collar-and-criminal-law/as-googles-waymo-expands-so-does-a-police-surveillance-tool [Online; accessed 19. Dec. 2023].

[85] Giulio Lovisotto, Henry Turner, Ivo Sluganovic, Martin Strohmeier, and Ivan Martinovic. 2021. SLAP: Improving Physical Adversarial Examples with Short-Lived Adversarial Perturbations. In *30th USENIX Security Symposium (USENIX Security)*. USENIX, Berkeley, CA, USA, 1865–1882.

[86] Mulong Luo, Andrew C Myers, and G Edward Suh. 2020. Stealthy Tracking of Autonomous Vehicles with Cache Side Channels. In *29th USENIX Security Symposium (USENIX Security)*. USENIX, Berkeley, CA, USA, 859–876.

[87] Steven Macenski, Tully Foote, Brian Gerkey, Chris Lalancette, and William Woodall. 2022. Robot Operating System 2: Design, Architecture, and Uses In The Wild. *Science Robotics* 7, 66 (2022), eabm6074.

[88] Yanmao Man, Raymond Muller, Ming Li, Z. Berkay Celik, and Ryan Gerdes. 2023. That Person Moves Like a Car: Misclassification Attack Detection for Autonomous Systems Using Spatiotemporal Consistency. In *32nd USENIX Security Symposium (USENIX Security)*. USENIX, Berkeley, CA, USA, 6929–6946.

[89] Jarrod R. McClean, Christopher Stull, Charles Farrar, and David Mascareñas. 2013. A Preliminary Cyber-Physical Security Assessment of the Robot Operating System (ROS). In *Unmanned Systems Technology XV*, Vol. 8741. SPIE, Bellingham, WA, USA, 874110.

[90] Tripp Mickle, Cade Metz, and Yiwen Lu. 2023. *G.M.'s Cruise Moved Fast in the Driverless Race. It Got Ugly.* The New York Times. https://www.nytimes.com/2023/11/03/technology/cruise-general-motors-self-driving-cars.html

[91] Muhammad Jehanzeb Mirza, Cornelius Buerkle, Julio Jarquin, Michael Opitz, Fabian Oboril, Kay-Ulrich Scholl, and Horst Bischof. 2021. Robustness of Object Detectors in Degrading Weather Conditions. In *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*. IEEE, New York, NY, USA, 2719–2724.

[92] Shuojie Mo, Xiaofei Pei, and Chaoxian Wu. 2022. Safe Reinforcement Learning for Autonomous Vehicle Using Monte Carlo Tree Search. In *IEEE Transactions on Intelligent Transportation Systems*, Vol. 23. IEEE, New York, NY, USA, 6766–6773.

[93] Raymond Muller, Yanmao Man, Z. Berkay Celik, Ming Li, and Ryan Gerdes. 2022. Physical Hijacking Attacks against Object Trackers. In *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security*. ACM, New York, NY, USA, 2309–2322.

[94] OpenRTM-aist. 2023. OpenRTM-aist. https://www.openrtm.org/openrtm/. [Online; accessed 28-Dec-2023].

[95] NHTSA Standing General Order. 2021. *Standing General Order on Crash Reporting*. Standard. National Highway Traffic Safety Administration.

[96] Poojan Oza, Vishwanath A. Sindagi, Vibashan Vishnukumar Sharmini, and Vishal M. Patel. 2024. Unsupervised Domain Adaptation of Object Detectors: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 46 (2024), 4018-4040 pages.

[97] Matthew O'Kelly, Aman Sinha, Hongseok Namkoong, John Duchi, and Russ Tedrake. 2018. Scalable End-to-End Autonomous Vehicle Testing via Rare-event Simulation. In *Proceedings of the 32nd Conference on Neural Information Processing Systems (NeurIPS)*. Neural Information Processing Systems Foundation, San Diego, CA, USA, 12 pages.

[98] Ugo Pagallo. 2013. *The Laws of Robots Crimes, Contracts, and Torts*. Vol. 10. Springer, New York, NY, USA.

[99] Fei Pan, Sungsu Hur, Seokju Lee, Junsik Kim, and In So Kweon. 2022. ML-BPM: Multi-teacher Learning with Bidirectional Photometric Mixing for Open Compound Domain Adaptation in Semantic Segmentation. In *European Conference on Computer Vision (ECCV)*. Springer, New York, NY, USA, 236–251.

[100] Fei Pan, Inkyu Shin, Francois Rameau, Seokju Lee, and In So Kweon. 2020. Unsupervised Intra-domain Adaptation for Semantic Segmentation through Self-Supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, New York, NY, USA, 3764–3773.

[101] Fei Pan, Xu Yin, Seokju Lee, Axi Niu, Sungeui Yoon, and In So Kweon. 2024. MoDA: Leveraging Motion Priors from Videos for Advancing Unsupervised Domain Adaptation in Semantic Segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. IEEE, New York, NY, USA, 2649–2658.

[102] Yet Another Robot Platform. 2023. *YARP*. YARP. Retrieved December 28, 2023 from https://www.yarp.it/latest/

[103] Shiva Raj Pokhrel and Jinho Choi. 2020. A Decentralized Federated Learning Approach for Connected Autonomous Vehicles. In *Pro2020 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*. IEEE, New York, NY, USA, 6 pages.

[104] PX4. 2023. *PX4*. PX4. Retrieved December 28, 2023 from https://px4.io/

[105] Morgan Quigley, Ken Conley, Brian Gerkey, Josh Faust, Tully Foote, Jeremy Leibs, Rob Wheeler, and Andrew Y. Ng. 2009. ROS: an Open-Source Robot Operating System. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA) Workshop on Open Source Software*. IEEE, New York, NY, USA, 5.

[106] UN R155. 2021. *Uniform Provisions Concerning the Approval of Vehicles with Regards to Cyber Security and Cyber Security Management System*. Standard. United Nations.

[107] Aruni RoyChowdhury, Prithvijit Chakrabarty, Ashish Singh, SouYoung Jin, Huaizu Jiang, Liangliang Cao, and Erik Learned-Miller. 2019. Automatic Adaptation of Object Detectors to New Domains Using Self-Training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, New York, NY, USA, 780–790.

[108] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Fei-Fei Li. 2015. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision* 115 (2015), 211–252.

[109] Kuniaki Saito, Yoshitaka Ushiku, Tatsuya Harada, and Kate Saenko. 2019. Strong-Weak Distribution Alignment for Adaptive Object Detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, New York, NY, USA, 6956–6965.

[110] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. 2018. Semantic Foggy Scene Understanding with Synthetic Data. *International Journal of Computer Vision* 126 (2018), 973–992.

[111] Erica Salvato, Gianfranco Fenu, Eric Medvet, and Felice Andrea Pellegrino. 2021. Crossing the Reality Gap: A Survey on Sim-to-Real Transferability of Robot Controllers in Reinforcement Learning. *IEEE Access* 9 (2021), 153171–153187.

[112] Takami Sato, Sri Hrushikesh Varma Bhupathiraju, Michael Clifford, Takeshi Sugawara, Qi Alfred Chen, and Sara Rampazzi. 2023. WIP: Infrared Laser Reflection Attack Against Traffic Sign Recognition Systems. In *Proceedings of the Inaugural International Symposium on Vehicle Security & Privacy (VehicleSec '23)*. ISOC, Reston, VA, USA, 5 pages.

[113] Takami Sato, Sri Hrushikesh Varma Bhupathiraju, Michael Clifford, Takeshi Sugawara, Qi Alfred Chen, and Sara Rampazzi. 2024. Invisible Reflections: Leveraging Infrared Laser Reflections to Target Traffic Sign Perception. In *Proceedings of the Network and Distributed System Security Symposium (NDSS '24)*. ISOC, Reston, VA, USA, 18.

[114] Takami Sato, Junjie Shen, Ningfei Wang, Yunhan Jia, Xue Lin, and Qi Alfred Chen. 2021. Dirty Road Can Attack: Security of Deep Learning Based Automated Lane Centering Under Physical-World Attack. In *Proceedings of the 30th USENIX Security Symposium (USENIX Security 21)*. USENIX, Berkeley, CA, USA, 3309–3326.

[115] Scale. 2023. *Scale*. Scale. Retrieved December 28, 2023 from https://scale.com/blog/afm1

[116] Umar Shakir. 2023. *Robotaxis are Driving on Thin Ice*. The Verge. Retrieved December 22, 2023 from https://www.theverge.com/2023/8/18/23837217/cruise-robotaxi-driverless-crash-fire-truck-san-francisco

[117] Shai Shalev-Shwartz, Shaked Shammah, and Amnon Shashua. 2017. *On a Formal Model of Safe and Scalable Self-Driving Cars*. Mobileye. arXiv:1708.06374 [cs.RO]

[118] Hocheol Shin, Dohyun Kim, Yujin Kwon, and Yongdae Kim. 2017. Illusion and Dazzle: Adversarial Optical Channel Exploits against Lidars for Automotive Applications. In *Proceedings of the Conference on Cryptographic Hardware and Embedded Systems 2017 (CHES '17)*. Springer, New York, NY, USA, 21 pages.

[119] Vishwanath A. Sindagi, Poojan Oza, Rajeev Yasarla, and Vishal M. Patel. 2020. Prior-Based Domain Adaptive Object Detection for Hazy and Rainy Conditions. In *Proceedings of the European Conference on Computer Vision (ECCV)*. Springer, New York, NY, USA, 763–780.

[120] IEEE Vehicular Technology Society. 2022. *Assumptions in Safety-Related Models for Automated Driving Systems*. Standard. Institute of Electrical and Electronics Engineers. 1–59 pages.

[121] Sheikh M. Solaiman. 2017. Legal Personality of Robots, Corporations, Idols and Chimpanzees: a Quest for Legitimacy. *Artificial Intelligence and Law* 25 (2017), 155–179.

[122] Ruoyu Song, Muslum Ozgur Ozmen, Hyungsub Kim, Raymond Muller, Z. Berkay Celik, and Antonio Bianchi. 2023. Discovering Adversarial Driving Maneuvers against Autonomous Vehicles. In *Proceedings of the 32nd USENIX Security Symposium (USENIX Security 23)*. USENIX, Berkeley, CA, USA, 2957–2974.

[123] Takeshi Sugawara, Benjamin Cyr, Sara Rampazzi, Daniel Genkin, and Kevin Fu. 2020. Light Commands: Laser-Based Audio Injection Attacks on Voice-Controllable Systems. In *Proceedings of the 29th USENIX Security Symposium (USENIX Security 20)*. USENIX, Berkeley, CA, USA, 2631–2648.

[124] Jiachen Sun, Yulong Cao, Qi Alfred Chen, and Z. Morley Mao. 2020. Towards Robust LiDAR-Based Perception in Autonomous Driving: General Black-Box Adversarial Sensor Attack and Countermeasures. In *Proceedings of the 29th USENIX Security Symposium (USENIX Security 20)*. USENIX, Berkeley, CA, USA, 877–894.

[125] Pascalis Trentsios, Mario Wolf, and Detlef Gerhard. 2022. Overcoming the Sim-to-Real Gap in Autonomous Robots. *Procedia CIRP* 109 (2022), 287–292.

[126] Timothy Trippel, Ofir Weisse, Wenyuan Xu, Peter Honeyman, and Kevin Fu. 2017. WALNUT: Waging Doubt on the Integrity of MEMS Accelerometers with Acoustic Injection Attacks. In *Proceedings of the 2017 IEEE European symposium on security and privacy (EuroS&P)*. IEEE, New York, NY, USA, 3–18.

[127] James Tu, Mengye Ren, Ming Liang, Bin Yang, Richard Du, Frank Cheng, and Raquel Urtasun. 2020. Physically Realizable Adversarial Examples for LiDAR Object Detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2020 (CVPR '20)*. IEEE, New York, NY, USA, 13716–13725.

[128] David C Vladeck. 2014. Machines Without Principals: Liability Rules and Artificial Intelligence. *Wash. L. Rev.* 89 (2014), 117.

[129] Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno Olshausen, and Trevor Darrell. 2020. Tent: Fully Test-Time Adaptation by Entropy Minimization. In *Proceedings of the International Conference on Learning Representations (ICLR)*. ICLR, Appleton, WI, USA, 15 pages.

[130] Wei Wang, Yao Yao, Xin Liu, Xiang Li, Pei Hao, and Ting Zhu. 2021. I Can See the Light: Attacks on Autonomous Vehicles Using Invisible Lights. In *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*. ACM, New York, NY, USA, 1930–1944.

[131] Waymo. 2023. *Waymo Safety*. Waymo. Retrieved December 22, 2023 from https://waymo.com/safety/

[132] Wayve. 2023. *Wayve*. Wayve. Retrieved December 28, 2023 from https://wayve.ai/thinking/lingo-natural-language-autonomous-driving/

[133] Xinshuo Weng, Boris Ivanovic, Kris Kitani, and Marco Pavone. 2022. Whose Track Is It Anyway? Improving Robustness to Tracking Errors with Affinity-based Trajectory Prediction. In *Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, New York, NY, USA, 6563–6572.

[134] William H. Widen and Philip Koopman. 2022. Autonomous Vehicle Regulation & Trust: The Impact of Failures to Comply with Standards. *UCLA JL & Tech.* 27 (2022), 169.

[135] Hai Wu, Wenkai Han, Chenglu Wen, Xin Li, and Cheng Wang. 2021. 3D Multi-Object Tracking in Point Clouds Based on Prediction Confidence-Guided Data Association. *IEEE Transactions on Intelligent Transportation Systems* 23, 6 (2021), 5668–5677.

[136] Zuxuan Wu, Ser-Nam Lim, Larry S Davis, and Tom Goldstein. 2020. Making an Invisibility Cloak: Real World Adversarial Attacks on Object Detectors. In *Proceedings of the European Conference on Computer Vision (ECCV)*. Springer, New York, NY, USA, 17 pages.

[137] Qifan Xiao, Xudong Pan, Yifan Lu, Mi Zhang, Jiarun Dai, and Min Yang. 2023. Exorcising "Wraith": Protecting LiDAR-based Object Detector in Automated Driving System from Appearing Attacks. In *Proceedings of the 32nd USENIX Security Symposium (USENIX Security 23)*. USENIX, Berkeley, CA, USA, 2939–2956.

[138] Chulin Xie, Zhong Cao, Yunhui Long, Diange Yang, Ding Zhao, and Bo Li. 2019. *Privacy of Autonomous Vehicles: Risks, Protection Methods, and Future Directions*. University of Illinois Urbana-Champaign. arXiv:2209.04022 [cs.AI]

[139] Yun Xing, Dayan Guan, Jiaxing Huang, and Shijian Lu. 2022. Domain Adaptive Video Segmentation via Temporal Pseudo Supervision. In *Proceedings of the European Conference on Computer Vision (ECCV)*. Springer, New York, NY, USA, 621–639.

[140] Kaidi Xu, Gaoyuan Zhang, Sijia Liu, Quanfu Fan, Mengshu Sun, Hongge Chen, Pin-Yu Chen, Yanzhi Wang, and Xue Lin. 2020. Adversarial T-Shirt! Evading Person Detectors in a Physical World. In *Proceedings of the European Conference on Computer Vision (ECCV)*. Springer, New York, NY, USA, 665–681.

[141] Chen Yan, Wenyuan Xu, and Jianhao Liu. 2016. Can You Trust Autonomous Vehicles: Contactless Attacks Against Sensors of Self-Driving Vehicle. *Def Con* 24, 8 (2016), 109.

[142] Kaichen Yang, Tzungyu Tsai, Honggang Yu, Max Panoff, Tsung-Yi Ho, and Yier Jin. 2021. Robust Roadside Physical Adversarial Attack Against Deep Learning in Lidar Perception Modules. In *Proceedings of the ACM Asia Conference on Computer and Communications Security (ASIA CCS)*. ACM, New York, NY, USA, 349–362.

[143] Chengzeng You, Zhongyuan Hau, and Soteris Demetriou. 2021. Temporal Consistency Checks to Detect LiDAR Spoofing Attacks on Autonomous Vehicle Perception. In *Proceedings of the Workshop on Security and Privacy for Mobile AI (MAISP)*. ACM, New York, NY, USA, 13–18.

[144] Fisher Yu, Haofeng Chen, Xin Wang, Wenqi Xian, Yingying Chen, Fangchen Liu, Vashisht Madhavan, and Trevor Darrell. 2020. Bdd100k: A Diverse Driving Dataset for Heterogeneous Multitask Learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, New York, NY, USA, 2636–2645.

[145] Jun Zhang, Lei Pan, Qing-Long Han, Chao Chen, Sheng Wen, and Yang Xiang. 2021. Deep Learning Based Attack Detection for Cyber-Physical System Cybersecurity: A Survey. *IEEE/CAA Journal of Automatica Sinica* 9, 3 (2021), 377–391.

[146] Louise Zhang. 2023. *Cruise's Safety Record Over 1 Million Driverless Miles*. Cruise. Retrieved December 22, 2023 from https://getcruise.com/news/blog/2023/cruises-safety-record-over-one-million-driverless-miles/

[147] Ganlong Zhao, Guanbin Li, Ruijia Xu, and Liang Lin. 2020. Collaborative Training Between Region Proposal Localization and Classification for Domain Adaptive Object Detection. In *Proceedings of the European Conference on Computer Vision (ECCV)*. Springer, New York, NY, USA, 86–102.

[148] Xiaopei Zhu, Zhanhao Hu, Siyuan Huang, Jianmin Li, and Xiaolin Hu. 2022. Infrared Invisible Clothing: Hiding from Infrared Detectors at Multiple Angles in Real World. In *Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR '22)*. IEEE, New York, NY, USA, 13317–13326.

[149] Yi Zhu, Chenglin Miao, Foad Hajiaghajani, Mengdi Huai, Lu Su, and Chunming Qiao. 2021. Adversarial Attacks Against LiDAR Semantic Segmentation in Autonomous Driving. In *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems*. ACM, New York, NY, USA, 329–342.

[150] Yi Zhu, Chenglin Miao, Tianhang Zheng, Foad Hajiaghajani, Lu Su, and Chunming Qiao. 2021. Can We Use Arbitrary Objects to Attack LiDAR Perception in Autonomous Driving?. In *2021 ACM SIGSAC Conference on Computer and Communications Security (CCS '21)*. ACM, New York, NY, USA, 1945–1960.