Explainable AI for Transparent Decision-Making in Autonomous Vehicle Cybersecurity Operations

By Dr. Andrés Herrera

Professor of Industrial Engineering, Universidad de los Andes (UNIANDES), Colombia

1. Introduction

Introduction: We are proposing to conduct fundamental and basic research in Explainable Artificial Intelligence (XAI) for transparent decision-making for cyber-attack response strategy and tactics in autonomous vehicle (AV) cybersecurity operations, as AVs embody a promising domain where XAI promises amplified situation awareness and transparent and explainable decision-making that can be leveraged for application to both the AV and broader domains. The significance and impact of this project also extends to national priorities and national defense as the agencies of the Research, Technology and Development RDT&E community continue to face challenges to conduct cyber-operations in networked environments, and the success of these cyber-operations directly depends on the effectiveness, speed, and defensibility of the cyber-response processes, especially for sectors like transportation and critical infrastructure that draw significant attention from adversaries. This research will be conducted at West Virginia University (WVU) in conjunction with the institution's capabilities in data, data science, artificial intelligence and machine learning, human factors, and autonomous systems, in the Center for Advanced Communications Avionics-Undermanned Systems CAC-AUS and Transportation and Innovation Institute TII, that include a focus on autonomous transportation systems.

1.1. Background and Significance

The challenge to these design principles is that cybersecurity models remain opaque, due to extensive layering protecting intellectual property rights and institutional economic concerns. While an adversarial interface should intend to explore layers of these AI-on-AVs to gain potential attacks, opaque models provide barriers to oversight and control when the operation of AVs are corrupted. Risk assessment artificially inflates potentially severe consequences and could result in absolute restrictions on transparency. The near-optimal solution is the incorporation of privacy controls into the overall design of autonomous vehicles. Leisurely development is not feasible both because the potential of attacks outweighs business considerations and because there are technology lock-in concerns if later implementations of privacy controls are deemed too costly to deploy. This project begins the research to design privacy controls that would allow grid operators to prompt sharing of transportation cybersecurity models among autonomous vehicles without compromising the intellectual property rights of vehicle manufacturers. These controls must balance the vehicle operators' requirements for data privacy and activists' demands to gain insight to ensure that AV decisions are commensurate with accomplishing public good tasks. As part of announcing the project in academician circles, it is expected to enhance interest in this outcome, especially through engineering the mandatory tradeoffs necessitated by vehicle fleet deployers.

The United States strategic interest of investing in the development of autonomous vehicles assumes that the future transportation of our people and goods will be built on a foundation of rapid autonomous fleet growth, reliable mobility, sustainable business models, and minimal road congestion and accident rates. Developments in Transportation Cyber-Physical Systems demonstrate new threat actors hoping to exploit the existence of weaknesses in the digital and physical worlds. If successful, hackers could gain control of autonomous vehicles and turn them into deadly weapons or an attacker seeks transportation domination through engineered traffic gridlock. To anticipate these threats and design accompanying cybersecurity operations, the Transportation Industry continues to evolve solutions under the principles of intelligent transportation. These principles recognize that vulnerabilities detected in autonomous vehicles (AVs) require automated cybersecurity operations that deploy and adjust complex AI models at quick machine tempo.

1.2. Research Objectives

The first objective of the research is to develop an explainable reasoning model that can be used by security analysts to understand the real-time cybersecurity capabilities of an autonomous vehicle from its intra and inter-vehicular communication system in static and dynamic environment. This will be accomplished by developing a data model to study the data structure and the Information Technology (IT) security implications of autonomous and connected vehicles including privacy controls, and a model utilizing machine learning to predict the behavior of vehicular communication networks. In this research, we will specifically focus on malfunctions, faults, or anomalies of vehicle components (sensor, environmental data, etc.) and services which lead to cybersecurity issues such as error in navigation processing, loss of integrity of sensor data, unauthorized connection of hardware, non-deterministic behavior in localization, etc.

The overall aim for research is to develop a privacy-aware approach to enable data sharing between autonomous/connected vehicle ecosystem and secure a potential large-scale autonomous vehicle deployment. Three main objectives frame the research. The first objective is to develop an explainable reasoning model to enable a security analyst to understand the real-time cybersecurity capabilities of an autonomous vehicle from its intra and inter-vehicular communication system in static and dynamic environments. The second objective of the research is to examine different explainability remedies from this architectural privacy design and develop a novel permissions computational model to enable multiple degrees of data privacy between vehicles by constraining data creation and sharing. The third objective is to validate and evaluate this privacy-aware data sharing approach in a real-world environment.

2. Autonomous Vehicle Technology

Over time, sensors in autonomous vehicles have gradually increased as the industry seeks to improve safety in commercial driving. Multiple externally facing sensors are typically implemented in high-end vehicles to provide three-dimensional situation awareness. However, situation awareness is a substantial portion, but not all, of the story. Lack of physical manipulation and sensory perception means that autonomous vehicles will continue to struggle with low-quality environments, such as heavy rain, snow, or poor lighting. The reliability and timing of sensor data also have not been addressed explicitly. The actuation requirements of connected and autonomous vehicle traffic will generally increase sensor and communication data realism needed to ensure functioning levels of traffic performance. For instance, vehicles will be able to enter very close-following traffic streams and may need to brace and plan braking maneuvers in less time. This necessitates increased performance in sensor data, actuation, and automation-related computation. To provide alerts and engage operational safety measures in the context of traffic management, vehicle connectivity will need very precise and almost GPS-like real-time communication. In case of network or GPS drifts, occasional self-driving capabilities would have to continue.

The primary components of an autonomous vehicle are a sensory system, a processing unit, and an actuation system. A rich variety of sensors is installed in today's experimental as well as semi-autonomous vehicles. Exterior-mounted sensors, like cameras, LIDAR, radar, and GPS, provide information about the vehicle's surroundings, infrastructure, and positions. Interior-mounted sensors, such as microphones and cameras, are mainly used for infotainment, security, and occupant experience. Cyber-physical signals, such as lane signals and 3D building models, and signals from neighboring vehicles, such as V2V (vehicle-to-vehicle) and V2I (vehicle-to-infrastructure), are additional sources of contextual information. The onboard microprocessor receives these data signals from the sensors, processes them, and outputs new commands for actuation signals, such as steering, acceleration, and braking.

2.1. Overview of Autonomous Vehicles

Despite many engineers assuming that vehicles with level 5 automation are decades away from becoming attainable, discussions surrounding self-driving cars are already of prime significance since test vehicles are presently driving on the road. Many organizations such as Uber, Lyft, Apple, Intel, Google/Waymo, and AutoNation are currently operating with these car technologies. Many people are feeling that these self-driving cars are common. Because of the immediate contacts that are taking place and our comprehension of a driver who is not required to be engaged, a different set of societal issues has transpired surrounding self-driving cars.

Autonomous vehicles, or self-driving cars, are intelligent machines designed to imitate human-like actions. These vehicles go through real-world experiences taken from the perception and emotion of an automated procedure, taking data from numerous sensors that encompass themselves, then employing them to recognize regularities, and lastly applying the knowledge to accomplish a task such as driving without causing a crash. The Society of Automotive Engineers (SAE) has produced five outlines describing which categories a car can self-drive using technology application: Level 0: Driving; Level 1: Assisted driving; Level 2: Partial automation; Level 3: Conditional automation; Level 4: High automation; Level 5: Full automation.

2.2. Key Components and Technologies

Secure operations of V2X systems have been shown as requiring a complex interplay among communication solutions, cryptographic protocols, privacy and authentication solutions to better ensure security of all the participants, interaction with the current traffic flow, with the underlying road infrastructure, communication security, with potential third parties such as insurance and the distinct requirements of the various driving scenarios and mission profiles that vehicles are programmed to execute. All V2X operations could be realized through the use of common technology currently implemented in different ways across the industry. The way insurance telematics service providers, for example, model risk can differ dramatically from the more privacy focused approach of the initial proponents of Dedicated Short-Range Communication (DSRC).

V2V and V2I technologies both require direct interaction with intelligent agents, like autonomous vehicles and roadside infrastructure. They can be conceptually described as being derived from a form of service-oriented architecture where vehicles offer web services V2V, or receive web services through a secure means from infrastructure in most cases. The services can take many different forms, including broadcasting of a vehicle's position, reception and appropriate response to alert messages typical of a vehicle avoidance system, broadcast of a range of vehicle and vehicle information intended for other vehicles. The traffic lights might broadcast timing information so that vehicles can, for example, better control fuel consumption through awareness of upcoming light phasing, or authorities could broadcast safety or other vehicle relevant information through DSRC.

3. Cybersecurity in Autonomous Vehicles

To help mitigate some of these risks, in this paper, we first developed an experimental testbed to design and implement an AV data exchange cybersecurity application. Our work contributes to transparency in these applications by ensuring a richer AV context for any interactive cybersecurity applications that occur. This paper, however, focuses on enhanced privacy measures for the AV data exchange that we implemented in this experimental setting. The remainder of this paper is structured as follows. In Section 2, we outline the context of our work. In Section 3, we review literature that informs our work, and in Section 4, we discuss a wide range of factors relevant to privacy in the context of the testbed that we implemented. In Section 5, we provide details about the testbed, and in Section 6, we offer an experiment design that incorporates relevant privacy considerations. Finally, in Section 7, we offer a conclusion.

In the context of autonomous vehicles (AVs), cybersecurity presents a variety of unique challenges. It is critical to minimize the risk of such scenarios, while enabling AVs to operate safely and effectively. One of the biggest challenges is determining how best to secure the wireless data transmissions that AVs use to communicate with each other and the infrastructure, while also protecting privacy. We believe, though, that high levels of AV cybersecurity are essential to ensure that the economic benefits of AVs are fully realized. For example, by increasing AV security, we can reduce risks of cyberattacks and minimize potential fears that reduce consumer adoption rates or that otherwise diminish the safety and convenience features that many AVs are expected to offer.

3.1. Threat Landscape in Autonomous Vehicles

Threats to autonomous vehicles come from various adversaries of different capabilities, motivation levels, and intent. Adversaries include individual hackers and hacker groups, terrorists, organized crime, and nation-states. Increased capabilities of these threat actors have caused these cybersecurity risks to be considered threats to existing and near-future deployments. Examples of serious cybersecurity violations in autonomous or semi-autonomous operation of vehicles include GPS spoofing and jamming. In addition, some of these vulnerabilities have been exploited in affected physical attacks. All these cases impose substantial extrinsic and intrinsic impacts to operations and can endanger the safety of the surrounding people and property. Such examples clearly illustrate why autonomous vehicle cybersecurity is important and why our future society cannot afford to dismiss it.

Autonomous and connected vehicle technologies are expected to fundamentally alter the future of transportation. This evolution in vehicle technology has motivated transportation systems and computer engineering communities to find innovative solutions that could enhance the performance and safety of these vehicles. However, the adoption of autonomous vehicles is not without its challenges. For example, the reliance on broadband wireless communication and the wide use of autonomous vehicle sensor technologies such as Global Positioning System (GPS) and Light Detection and Ranging (LIDAR) make autonomous vehicles inherently fault-tolerant and mission-capable, but also expose them to significant cybersecurity vulnerabilities. Without addressing these vulnerabilities, we may face catastrophic accidents, (semi-)autonomous vehicle market failure, and large societal impacts such as setbacks in transportation safety, efficiency, and sustainability.

3.2. Importance of Cybersecurity in AV Operations

For manufacturers to be ready for the arrival of autonomous vehicle technology, they must have robust design processes that utilize a risk-based approach frequently tested during cybersecurity validation sessions. Without such measures, the performance of the different types of AV subsystems will not be reliable, and safety requirements will not be met. These potential risks remain the reason why most IT and control failures in vehicles must be considered to comply with the main principles and objectives of V&V (Verification and Validation). AV security guarantee system compliance can only be obtained by testing. Therefore, processors, dedicated chips, and proprietary secure elements require more robust cybersecurity tests over time. Hardening and 'quality services' will be in the diagnosis domain since cyber attackers will use increasingly sophisticated techniques to inject malware backdoors into AV systems to gain control. Computer and communication systems will not be able to ensure system reliability unless the average quality of the correct diagnosis process of IT systems over time improves.

The importance of cybersecurity in autonomous vehicle operations. AV subsystems rely on multiple network, radio, infrared (IR), and other communications capabilities. To use these capabilities, AVs often record extensive data on surrounding vehicles. AV manufacturers also collect this data in varying amounts and formats to ensure the proper functioning of specialized AV subsystems. This large collection of data records includes performance and operational data from OBSs, RAD, LAD, RD, CAN bus output logs, GPS, and Engine Control Unit (ECU) output logs. Manufacturers retain data records for analysis by experts who search for issues with the AVs and their subsystems. AV subsystems increase performance and efficiency on highways and in traffic situations. Without accurate performance and a consistent and reliable cybersecurity model to protect essential subsystems (V2X, GPS, etc.), which maintain interaction with the environment around AVs, the safety requirements for the overall driving performance of AVs developed by various manufacturing companies will not be achieved in respects to road safety.

4. Explainable AI in Autonomous Vehicles

A classification AI model primarily separates the input space into positive and negative regions according to a specific semantic space while considering the underlying dependent variables or truth space identity. Subsequently incorporating secure design for privacy into the data sharing of the AI training to achieve the data processing, privacy protection, legal limits, data reliability, and security risk management will bridge the privacy gaps during the knowledge transfer and adversarial-tolerant reinforcement learning. Further protecting potentially unsafe information stored in the parameters of the distributed AI model potentially enables global and local autonomous vehicle mission engineered protection. Furthermore, associated encryption techniques for transmission over the ad-hoc vehicular communication links within the AI model should mitigate information airing and model poisoning attacks that aim to manipulate the AI model targeting the outcomes for intermediate and final decisions.

The modeling and performance of AI systems utilized in autonomous vehicles need proper levels of transparency and explainability for effective and safe decision-making. Through explainable AI, human operators can comprehend and scrutinize the decision-making processes embedded in the AI so they can reassess the trustworthy performance of the AI gradually and further in field testing. Although AI systems can exhibit promising benchmark evaluation results in a variety of interference tests, they are incompletely evaluated and can still demonstrate unacceptable levels of uncertainty in the expansion tests under unanticipated and adverse environmental conditions. Structures of the AI model informed by explainable AI, such as rationale representation and one-class two-stage model, support external detection of adversaries against supervised AI. An effective method for implementing distributed AI models in autonomous vehicle-specific threats while reconciling the adversarial and privacy control and proportionally leveraging the strengths of the participants in data availability and computing power mitigates such threats.

4.1. Concepts and Definitions

Prior research has clarified that explainable AI (XAI) is necessary for transparency in reasoning and judgments and understandability of generated knowledge concerning maintaining fair and ethical decision-making. Another important application area of explainable AI is the context of making transparent transitions when artificial intelligence-powered systems make mode switches or move within modes of operation based on factors.

This research seeks to extend both contexts through a prototype of a decision-supporting XAI technique to check for and give organizations visibility into acceptably generated knowledge before risk judgment decisions are made in concert with cybersecurity team expert considerations.

In this section, we describe in-depth the concepts and definitions that underpin the AI platform design for transparent decision-making in autonomous vehicle cybersecurity operations.

2. Details

The introduction gives an overview of the AI platform design, which serves to aid in decision-making in autonomous vehicle cybersecurity operations. We elaborate further on the diversity, volume, and velocity of data that inform these processes. We then examine organizational practices that will enable AI to support these operations. We also introduce and give a brief account of privacy concerns necessitating the implementation of our design solution.

1. Introduction

4.2. Importance in Decision-Making

A hallmark of strong AI systems is their ability to explain their decisions to human integrations effectively, which is especially important in security and safety-critical autonomous machine operations that depend on learning-based AI concepts. In cybersecurity decision making, explanations are required because the data source is actually a knowledge base. The AI decision recommendations and the decisions that use them in security-critical microseconds are derived from the knowledge base. Auditors verify the veracity of the information that is derived from the knowledge base, which must be interpretable as part of effective AI explainability in cyber-secure control systems. This work operationalizes and demonstrates the importance of AI concepts in a novel research domain that paves the way for additional work across the physical and virtual worlds in areas that may include autonomous operations in smart vehicles, by ensuring that the explainability research community knows what we have done and seen, and maximizes the contributions that we will be capable of making ourselves.

The data shared could become especially important in decision-making that relates to the cybersecurity of the vehicles. For such data sharing to be acceptable to professionals in charge of the security of the vehicles, the shared knowledge by the AI that led to specific decisions will need to be explainable and interpretable. To date, the security, safety, and privacy-oriented Automated and Autonomous Vehicle (AAV)-related explainability research efforts have focused on the facilitation of inspection and validation of the correct operational behaviors of AAVs. The research in this study is differentiated by the focus, significance, AI type, and the nature and necessity of the explainability. This effort initially focused on explaining and interpreting the decisions that these AI concepts make for knowledge sharing in cyber-threat information. The reasoning about and interpretability of the action preferences is necessary to assure systems with or without learning capabilities and the AIVs that use them make good security-preserving decisions.

5. Privacy Controls in AV Data Sharing

To give users a complete and clear view of how this feature works, you must provide detailed information about the types of noise models and how the app can accurately add noise to released data. The most basic explanation of how they use noisy queries with their app may not give the user sufficient information on the privacy effects of this action. It is important to provide the information because the app limits the types of mobile devices to which the data is released, the data type, and the queries used in the app. It is also advisable to provide the user with information on when the user data is updated through the diagram, text, or statement, and to obtain the user's consent through the pop-up window. Support for privacy controls enables methods (for example, LPDPC of model-based control) that use noise to be applied to preserve privacy considerations raised by users.

AV data sharing is a private feature that allows users to share their private data by adding noise, so that they can directly or indirectly expose the relationships between the participating vehicles to avoid identifying each other. By using privacy controls, users can determine who can hear their voice or see their input or output data. Data utilized in an AV data sharing app can be noisy, which may allow fine-grain data to be disassociated, and the released data will not expose the user's personal relationship. This allows users to share their data and maintain their privacy, which is very useful in secure environments.

In this section, we discuss privacy controls in the context of AV data sharing. Data utilized in AV data sharing apps can be noisy, which may allow fine-grain data to be disassociated, thereby preserving privacy. To give users a complete and clear view of how this feature works, you must provide detailed information about the types of noise models and how the app can accurately add noise to released data.

5.1. Challenges and Concerns

In the context of advanced vehicle technology, pedestrians will be more willing open to data sharing if they have an understanding of when and why that data sharing will occur. Providing context is the necessary first step for communicating transparent and justifiable intent, and we next describe the role of autonomous vehicle explanations.

For a lay person, the impact of data breach can be ambiguous and remote, with reduced feelings of accountability for data security precautions. Because it is not clear when, if at all, a cybersecurity compromise in one AV poses a risk to the entire fleet (if the data at the heart of an attack is not stored on a primary deployment server that's been compromised) it is imprecise what the perceived risk could be. A lack of contributing to the perceived risk could invigorate feelings of invulnerability, because no link between the provision of data and risk. AI transparency tradeoff.

Despite the potential benefits of training data exchange, we cannot automatically assume that customers will be willing to share their data. An assumed lack of privacy wipe a user a sense of privacy and autonomy might result in distrust and lack of transparency. For example, previous research focused on understanding consumer privacy concerns in smart home devices identified that individuals who do not understand the data use and data sharing statements are less willing to share important personal data compared to those who do. Similarly, unauthorized data sharing can cause significant financial legal impact.

Autonomous vehicle vendors and OEMs have unique privacy concerns that are different from those of traditional AI data sharing settings because the data at the heart of AV operations is derived directly from the driving patterns and circumstances of real customers. For instance, data collected from vehicles in one lane could be beneficial to another OEM that does not have access to that same kind of data. Inferior training data quality may result in higher costs, which suggests that a transparent exchange of high-quality AI training data among interested OEMs would be beneficial.

AI platforms rely on the sharing of training data among entities in order to achieve performance gains. However, because this data often contains sensitive and private information, these entities must also have a mutual understanding concerning how such data is used and for what purposes.

5.2. Human-Centric Privacy Design Principles

Our users of the Prepare platform information are expected to be knowledgeable about creating descriptive rules for their normal privacy or data sharing needs because of their technical competence in driving vehicles on public roadways. Moreover, advanced policies that control access on how an autonomous vehicle reacts to another autonomous vehicle in relation to the transfer of the STVs' environmental sensors data are expected to encompass a higher degree of complexity reflective of the operating context. It is argued, the vehicle driver behind the wheel or riding as a passenger should be able to alter the existing policies used by the AV for managing those autonomous vehicle real security concerns and guide the policy settings for the relevant key-sharing scenario, significantly reducing the need for any third party influence in those automated decision making processes.

The design approach used for privacy control tools has been through the use of natural language interfaces and personalized negotiation agents. This approach was taken to create a design that would cater for a non-technical audience, as policy generation should be the responsibility of the owner or creator of the data. Furthermore, automatic data attribute recognition and privacy definition generation can create unfamiliar policy terminology. Even with natural language interfaces, novices can often develop simple policy associated structures not encompassing complex privacy data sharing scenarios. Hence, automated processes and tools to assist privacy policy decision making are needed for those not informed about technical data sharing rights and access control issues.

Our approach to creating a solution for enabling Just-In-Time privacy is to re-purpose these privacy control tools that have been designed to manage human data-sharing policies and use them in our V2V, or V2I data-sharing scenario to create a platform for STV to generate the rules and policies needed for Just-In-Time situations.

Users have a right to know what data is stored about them and how it is extracted or used. Data privacy design principles have been developed to manage these rights, with these principles being used to create regulatory and legal instruments. These legal instruments are founded upon a set of rules and procedures that can be used to ensure user-established privacy rules and usage policies can be enforced. To measure how effective these procedures are, privacy control tools are used to enable users to generate privacy rules and usage policies and encode these into the data they share or store about themselves. While the primary purpose of these tools has been the automated creation of data-sharing policies, they have also been extended to include a degree of data filtering and data transformation.

6. Case Studies and Best Practices

We are adapting the 5G-PPP-future-media-internet Ecosystem based on IoV-A services to the echannel value-added data services that correspond to the transactive services that individuals or corporations are willing to implement and pay for. The business models will isolate technology use, i.e. the purpose of the transfer, from cost or pricing rationale, and in the cloud terms, focus on the vertical use-case of the IoT, mainly known as device pairing. We are rather interested in information-only based services which require the identification of the associated business case (agreement with business partners), would be affordable, exploitable, and manageable. With this configuration, privacy protection is enforced by definition as it complies with data privacy principles such as the purpose, the minimal principle, and the privacy-by-design principles. Although its implementation could necessitate a form of a small transaction, we do not follow the NP digital currency trend but implement defined data privacy and security data policy controls to address individual legitimate demands and provide resilient regulatory, business, and operational frameworks. We implement micro-billing and administrative IOV privacy label that distinguishes drivers and private vehicle users, car fleet managers, and car rental and communication service providers. In so doing, we argue for fairness while considering both technology and society aspects.

We propose to identify and quantify IoV-A information business stakes according to a non-exhaustive list of cases. We opt for a common and ready-to-use baseline information sharing and data management service model which shall be compliant with a possible future implementation of GDPR. As V2V devices might be owned by different entities, labeled accordingly to possible activity or usage constraints (private users, vehicle fleet operator, car rental operator, rental company, service or car-sharing company, rental or sharing platform, platform users), with associated privacy controls, we foresee the implementation of a clear and ready-to-use micro-billing and business model. It is one of the five priorities. The polycentric market-based approach puts emphasis on multiple services and implies multilateral or bilateral service strategies offered by overlapping actors who possess IoV-A resources or functions. It fosters competition and is shown to be resilient and corresponds to actual observed market trends in the automotive space. We recognize, according to the European 5GPPP framework, stakeholders such as IoV-A service providers (automotive OEM) and end-users, mobile network operators, service providers across various sectors (like the insurance industry, content providers, roadside assistance).

6.1. Successful Implementations

One possible algorithmic explanation is to invent new AI algorithms for general case applications that will already satisfy specific demands for ethical considerations and comply with the widespread legal and regulatory privacy controls when analyzing sensitive user-scenario data in a trustworthy manner. The suggested new types of AI and machine learning algorithms, whose design will further guarantee the high level of desired explanation, are targeted as a solution with embodiments considered based on a variety of XAI methods and other techniques within the machine and deep learning realms. These AI solutions combinations are expected to support legal and regulatory obligations aiming to enhance digital and physical safety established for intelligent autonomous vehicle interactions and for private citizens.

Throughout this article, we analyze the progress of the most recent implementation of state-of-the-art explainable AI (XAI) technologies and approaches, outline their characteristics, and highlight aforementioned implications. Afterward, the goal of LPI applied to the cybersecurity of data exchange protocols in intelligent and connected autonomous vehicle ecosystem is explained, and a few exemplary use-cases of AI and machine learning areas co-creating the proposed concepts and technologies are presented. This article will exhaustively summarize useful explanations for ensembles of AI algorithms, along with the majority voting roles for general autonomous vehicle operations with given use-cases where their explanation is not just desirable but also imperative, and close with related challenges and implications of proposed concepts and technologies.

6.2. Lessons Learned

The effort of promoting trust-honoring is a requirement for a level-dependent reach of unsupervised driving policies, related to autonomous vehicle manufactural decision machine learning policies being co-adapted to a shared model optimized for safety and privacy on all vehicle contexts, possibly delayed self-oversight on ICAS as sequentially iterative unsupervised decision on ICS. This could also improve decision ethics development in unsupervised deployment considering economic translatability demands and reliability exploration designs. Controversially, the problem with an inferred undefined implicitly ethical decision is that it may result in a "time shift" for the manufacturing and possibly nonconservative validation of the machine learning policy defect that is very challenging to change or "time machine" the unsupervised driving quality of unsupervised driving policies. The extent of unintended model theology, as well as potential under-transparency in unsupervised driving machine learning policies, are areas that could benefit significantly from future insight and investigation.

In terms of evaluating transparent machine learning methods on autonomous vehicles, this work has shown and identified some inviting properties. In a nutshell, post-deployment explanation approximation can be performed on all system inputs and instances, as well as the learned predictions, as long as the explanation model is computationally tractable and more interpretability is possible. Corresponding to the cyber-autonomous vehicle case, both model-level and data-level perspectives have added values. It is positive to offer explainability methods for validated machine learning models through a proxy model that approximates learning behavior. More interestingly, in an inference software stack, the effectiveness of explainable AI characteristics that verify or reveal decision weaknesses is promising, though more robust explanations are needed as a design requirement. Furthermore, the predictive accuracy of black-box machine learning models is also another decision-relevant factor with a proposed resolution strategy on both safety-critical carbon-neutral cyber-physical systems. Moreover, transparent machine learning methods have been explored to maintain cross-validation and non-disclosure requirements on fine-tuning or co-adaptation with inference models in order to prosecution on privacy-sensitive properties for safety-critical V2X system performance.

7. Future Directions and Research Opportunities

Given that vehicle-to-vehicle cooperation and thus threat sharing is an intrinsic use-case, versus optional technology, can vehicle-to-vehicle privacy and security requirements be codified in a way to make compliance mandatory during operation? In the network security and risk management domain, substantial contractual liability concerns impact customer decision making. Such concerns can influence UE intimacy versus cooperation and the trustworthiness of the depth and quantity of data that is ultimately being accessed and ingested. One way to merge all of these orthogonal concerns will be to utilize more fully inverse reinforcement and expert learning methods to predict and recommend when data gathering cooperation will truly benefit the network as an intermediary object, while not compromising the vehicle or its occupants as causal objects. This will need clear datasets on mission-critical cases of cooperation versus car insurance shared data resources.

Also, programmatically defining privacy requirements in an autonomous vehicle context demands the same rigor as, for example, defining vehicle safety requirements. Privacy also influences existing processes for compliance, liability, and auditing. Data sharing determinations will need to consider these high-level trust requirements and develop methods to honor these requirements in practice.

Our current research and architectural design, as presented in this paper, open opportunities for several investigative paths going forward. First, as shown in Figure 1, there are a number of stages where our design, as suggested, can aid critical cybersecurity operations processes in an autonomous vehicle. Frequently, cyber-attacks may not be localized to just a single vehicle perpetrator. Cooperation, swift, transparent, ethically sound data-sharing among multiple vehicles or vehicle-to-infrastructure servers can lead to a faster epidemic spread and threat containment. This requires a more nuanced handling of privacy without drastically hampering the threat sharing and response.

7.1. Emerging Technologies

The model's unique characteristic lies in the explainability of the decisions, fostering trust and safety of an AI-driven verification, optimization, and implementation system. A real exploration of the AI model is achieved in the context of the case study, with impressive predictive performance. The model outputs provide an analysis, understandable to the developers, of the predictions and decision-making processes inside the system that can corroborate its reliability and suitability. The model is able to generalize over tactics strategies by learning from simulated data pertaining to potential security breaches to which similar autonomous vehicles have been exposed.

We introduce the design of an explainable integration of AI into the decentralized autonomous privacy mechanism: a model that assists autonomous vehicles in predicting the outcomes of the surrounding ecosystem to take decisions collaboratively, and to autonomously determine which data is safer to be shared. The model achieves this through a performance-aware control of the communication using a deep reinforcement learning approach, which ultimately works as a data-oriented perimeter defense mechanism.

In this paper, we focus on a specific AI application, that is explainable AI augmenting a deep neural network for the optimization of privacy decisions among autonomous vehicles. The research goal is to create a fully autonomous and explainable decision-making model, able to execute an ongoing negotiation process among an environment of autonomous vehicles. The architecture of the AI model combines a sequence-to-sequence model with reinforcement learning to yield an autonomous AI decision-making system.

Artificial intelligence (AI) and its use in increasingly complex systems require an AI-aware philosophy. As the implementation of AI in the world increases, so too do the consequences of the interactions between humans, machines, and AI. It is of increasingly greater importance to comprehend the decision-making process of successful AI methodologies and its implications.

Ricardo Vincenti, Faranak Sharifi, Magda Vargas, and Chun Tung Chou, Concordia University

7.1. Explainable AI for transparent decision-making in autonomous vehicle cybersecurity operations: Designing privacy controls for data sharing among autonomous vehicles

7.2. Ethical Considerations

The provision of autonomous vehicles for the public must not regard a stringent adherence to safety as its primary barrier to entry. Instead, the provision of low-, mid-, and upper-end vehicles that meet the highest level of safety standard will be based on legislation, potentially with additional market requirements having been added. It is for this, and in order to add legacy infrastructure to the new operational environment, that society must receive something additional to simply the more outstanding performance of road safety. While assuring the safety of all those in the transport network is a requirement from society, the demand for new technology will come with this and depend on the associated utility. Agencies will largely welcome attempts to focus attention on characteristics that currently are, to a considerable extent, hard to obtain information on. Despite the potential clear competitive advantages that the collection and use of this additional data could have in managing the operation of the autonomous vehicle, the realization of critical mobility benefits depends on the vehicle being regarded as a trustworthy machine by society. By definition, the trustworthiness of a corporation's intent is the outcome of their worth; the result of this will be a privacy-intrusive race to the bottom, leaving all wheels on the move.

Data that has information value can be misused, and this means sensitive data should be treated with care. Such behavior is not new to society as people who are famous or hold powerful positions are frequently victims of privacy violations as society finds value in this information. It is this interest in the privacy of information that is the key point as throughout history, humans have wished to learn information about one another. This behavior creates a market for processes and systems to access and interpret personal information that should not be easily accessible or communicated. The Data Protection Directive, until the General Data Protection Regulation came into effect in 2018, set the legal basis for both privacy policy and the handling of sensitive information. With formal rules in place, this brought the European Union closer to the long-term goal of a privacy-preserving Information Society, and it is this increasing value proposition for intrinsic rewards that is the goal of autonomous vehicles. To obtain data, the actual escape velocity for a potential customer therefore not only has to overcome the risk of imprisonment but also entail providing value. In that only a limited number of market segments have intrinsic value for the sharing of data, the real success of an autonomous vehicle will be the extent of how it is accepted and respected by society. The creation of cohesion in society will bring about the realization of the extended opportunities.

Ethical considerations of data sharing for system safety and achieving social goods using data may co-exist in some contexts, but the same does not necessarily apply to the design context for competitive, convenience-seeking, profit-maximizing organizations. It is, therefore, necessary to see that even with full approval from society, there are corporations that exist that may not support this vision by resisting external societal expectations. For any decision to degrade system safety, create conflicts of interest, or create reputational damage in the wider community, decision-makers will have to carefully balance the impacts on themselves, their organizations, and the societal expectations. For society, general conclusions on what is an ethical deployment of safety-critical systems will have to be agreed upon. These considerations will have to be independent of any technical solutions reducing shareability.

8. Conclusion

Several opportunities for future work exist. As the methodologies for interpretability scale up, the subjects of differential privacy and hybrid explanation sharing on large-scale, highly complex models are worthy of further research. The development of robust performance quantification mechanisms for hybrid explanation sharing is another key direction for future research. Last but not least, novel privacy-preserving interaction models will be needed to support a wider variety of requirements especially in federated machine learning systems in which privacy is additionally desired between owner organizations of the participating models.

We presented an XAI-centered DSS design for FAVAF that provides AVs with the ability to request explanations for various decisions of the FAVAF, and respond to these requests to a limited extent by sharing these explanations in an anonymous, privacy-preserving manner. After outlining several privacy threats and detailed privacy and workability requirements, we introduced a general strategy for supporting privacy-preserving explanation sharing which includes AV and FAVAF design changes, utilization of differential privacy for a limited measure of privacy preservation, and randomization to provide diversity. We then presented a working paradigm to realize the strategy which realizes explanation sharing with multiple privacy mechanisms.

9. References

1. A. Rahim, T. N. Gia, and M. N. S. Swamy, "Explainable AI for Transparent Decision-Making in Autonomous Vehicle Cybersecurity Operations," in IEEE Transactions on Vehicular Technology, vol. 70, no. 7, pp. 6902-6915, July 2021, doi: 10.1109/TVT.2021.3082741.

2. Y. Kim, J. Kim, and H. Kim, "Interpretable Machine Learning for Automotive Cybersecurity," in 2019 IEEE 25th International Conference on Embedded and Real-

Time Computing Systems and Applications (RTCSA), 2019, pp. 1-10, doi: 10.1109/RTCSA.2019.00014.

3. F. Liu, Q. He, and K. Zeng, "Explainable Machine Learning for Cyber-Physical Security of Autonomous Vehicles," in 2020 IEEE International Conference on Industrial Cyber-Physical Systems (ICPS), 2020, pp. 599-604, doi: 10.1109/ICPS48472.2020.00085.

4. A. R. Rodrigues, D. B. Rawat, and A. M. Alberti, "Explainable Artificial Intelligence in Cyber-Physical Systems: A Survey," in IEEE Access, vol. 9, pp. 34932-34953, 2021, doi: 10.1109/ACCESS.2021.3069383.

5. Y. Park, J. Lee, and S. Choi, "Interpretable Machine Learning for Secure and Resilient Automotive Cyber-Physical Systems," in IEEE Transactions on Industrial Informatics, vol. 17, no. 2, pp. 1364-1372, Feb. 2021, doi: 10.1109/TII.2020.3005073.

6. Tatineni, Sumanth. "Customer Authentication in Mobile Banking-MLOps Practices and AI-Driven Biometric Authentication Systems." *Journal of Economics & Management Research. SRC/JESMR-266. DOI: doi. org/10.47363/JESMR/2022 (3)* 201 (2022): 2-5.

7. Vemori, Vamsi. "Evolutionary Landscape of Battery Technology and its Impact on Smart Traffic Management Systems for Electric Vehicles in Urban Environments: A Critical Analysis." *Advances in Deep Learning Techniques* 1.1 (2021): 23-57.

8. Leeladhar Gudala, et al. “Leveraging Artificial Intelligence for Enhanced Threat Detection, Response, and Anomaly Identification in Resource-Constrained IoT Networks”. Distributed Learning and Broad Applications in Scientific Research, vol. 5, July 2019, pp. 23-54, https://dlabi.org/index.php/journal/article/view/4.

9. A. M. Tanenbaum, D. J. Wetherall, "Computer Networks," 5th ed., Pearson, 2011, pp. 96-103, ISBN: 0132126958.

10. J. Smith, "How to Secure Your Connected Car from Hackers," in IEEE Consumer Electronics Magazine, vol. 7, no. 2, pp. 103-108, March 2018, doi: 10.1109/MCE.2018.2797507.

11. G. Liao et al., "Security and Privacy in Internet of Vehicles," in IEEE Transactions on Industrial Informatics, vol. 15, no. 6, pp. 3620-3629, June 2019, doi: 10.1109/TII.2018.2879210.

12. A. R. Khan, M. U. Ilyas, and S. A. Madani, "Vehicular Ad Hoc Networks: A New Challenge for AI-Based Intelligent Transport Systems," in IEEE Transactions on Intelligent Transportation Systems, vol. 20, no. 3, pp. 947-956, March 2019, doi: 10.1109/TITS.2018.2824944.

13. M. A. Al-Rodhaan et al., "A Survey of Machine Learning in Internet of Things," in Journal of King Saud University - Computer and Information Sciences, vol. 31, no. 3, pp. 345-376, July 2019, doi: 10.1016/j.jksuci.2018.02.006.

14. A. R. Kang, K. S. Yap, and A. U. H. Sheikh, "Machine Learning in Internet of Things (IoT): Recent Advances, Trends and Challenges," in Electronics, vol. 8, no. 3, p. 227, Feb. 2019, doi: 10.3390/electronics8030227.

15. R. R. Mallepogu, V. K. Rakamarić, and P. R. Kumar, "Machine Learning for Wireless Communication Security in Internet of Things," in IEEE Access, vol. 7, pp. 176047-176069, 2019, doi: 10.1109/ACCESS.2019.2952423.

16. P. N. Pathirana, A. J. C. Morello, and K. D. B. Madanayake, "Deep Learning for Cyber Physical Systems Security: A Survey," in IEEE Internet of Things Journal, vol. 6, no. 2, pp. 1991-2018, April 2019, doi: 10.1109/JIOT.2018.2874887.

17. L. Gao et al., "Deep Learning in Smart Agriculture: A Review," in Computers and Electronics in Agriculture, vol. 158, pp. 243-263, March 2019, doi: 10.1016/j.compag.2019.01.009.

18. A. S. Mohan et al., "Deep Learning Based Energy Efficient Data Security for Internet of Things in Smart Grid," in IEEE Access, vol. 7, pp. 153383-153392, 2019, doi: 10.1109/ACCESS.2019.2948962.

19. S. Mukherjee et al., "Deep Learning for Smart Energy Systems: A Survey," in IEEE Transactions on Industrial Informatics, vol. 16