Federated Learning for Privacy-Preserving Autonomous Vehicle Data Analysis

By Dr. Dzmitry Tsetseruk

Associate Professor of Computer Science, Belarusian National Technical University

1. Introduction

Moreover, the raw data that can be sensed, stored, and processed by autonomous vehicles are most likely to contain a lot of private attributes. Many of the personal behaviors such as long-term visiting habits, individual driving style, and demographic characteristics can be inferred from sensor data from autonomous vehicles. Potential adversaries to privacy can also exploit the vehicle data to recognize individuals, monitor their behaviors, and even damage their reputation [1]. Consequently, it is important to develop a mechanism that preserves the data privacy of the autonomous vehicles, as well as the privacy of inferred attributes by training machine learning models using the data.

The progress in self-driving and autonomous vehicles has brought some new security and privacy concerns that had previously been relatively unaddressed, since these technologies were not as pervasive. For example, data traversing from one point to another in the V2X ecosystem can be a privacy risk [2]. This risk becomes intensified as the vehicle is being driven in some remote location and has access to the local infrastructure over a potential malicious network which might be capable of eavesdropping and potentially participating in issuing wrong commands to the vehicle.

1.1. Background and Motivation

Car-makers and their software producers are not distributed clients for data management and model learning. Indeed, the accuracy and performance of car driving models also greatly modify driving models by deciding to adopt various solution methods via individual developers. A single FL kick can successfully assign the tribus achieved at a strikingly accurate model. When no FL protection is boarding swift or tiny by design for the future developments, especially when the gatherings are lacking diachronic data or the end-to-end learning tasks involve some privacy characteristics [3]. At last, while deploying a local trained model in a non-cooperative way, the plaintiff in a test bed bias becomes more important in deploying a local trained model to attack an opponent. As opposed to learning from singularly a fixed or steady state of model.

Federated learning (FL) is a defacto approach of modern machine learning to protect data privacy. It decouples the model training from the data individually owned by the multifarious parties by allowing the local model parameter updates to be uploaded from the parties for global model aggregation [4]. From then on, federated visual challenges like when local data in the traditional LTL paradigm is held by the entities, federated machine learning leverages the advantage of distributed learning frameworks, transforming the local data into the global model. The inherent nature of FL keeps this data inside the parties in which the data ha s been originally generated, thus supporting privacy. As a result, federated learning has been a hot spot for the consideration of the cloud computing or Internet of Things communities lately. In the FL context, a set of devices is used to analyze a set of autonomous car data sources in various privacy-preserving settings. In promptly withdrawing the risks associated with Chinese companies, a more convenient way of studying such data would be to randomly pick a company located in the United States, for example. Different from traditional machine learning algorithms, FL is employed as a defacto method to sequentially disperse the possible risk of linking data and to dramatically eliminate the danger of adversarial attack.

1.2. Research Objective

This chapter also suggests that new research directions on how to deploy FL in the automotive domain and, more in general, on cyber-physical systems. In such safety-critical domains, the architecture, the mobility model, and the network layer must be designed to guarantee (but also certify) the security and privacy of information and the robustness of control strategies. The Google's federated learning approach is to be in fact considered as a valid candidate to guarantee the indivual privacy of pedestrians and drivers when retraining deep learning models for predicting future events useful in scenarios, where anomaly detection or intention-aware control systems are required [5].

Privacy-Preserving Learning in the Autosafe Project [6] investigates mixted criticality systems from the point of view of a multilevel predictive mechanisms for intention prediction of cars. In the case study developed, developed swapped) is the advers busy road not on to the data is – and must §1.1. This data is analysed in real time with a TensorFlow deep learning model composed by Long Short-Term Memory (LSTM) layers. As shown in Figure 1, the model is re-trained using a privacy-preserving learning approach. This contribution will provide an overview on a privacy preserving deep learning solution for automotive data analysis and control systems.

1.3. Scope and Organization of the Work

Federated learning (FL) decouples the process of training machine learning models from the need of centralizing raw data on a single machine or at the cloud, by distributively training the model and aggregating the model updates locally at the end users' devices [5]. By engaging data owners to collaboratively generate gradient updates, FL essentially preserves the privacy of raw data. The adoption of FL in autonomous vehicle (AV) data analysis is a new research trend – it enables effective utilization of the massive sensor data generated by AVs, while protecting data privacy at the same time. Each model update is aggregated in a trusted third-party environment. For safety-critical systems such as automated vehicles, it has to be ensured that malicious behaviour, tainted data or inattentive clients do not compromise model integrity. Furthermore, when the vehicle re-enters the learning environment, the model should be able to adapt to changing operational situation and data characteristics. After pushing learning-based driver assistance systems and automated vehicle technologies firmly into development and pre-production status, the automotive industry aims to adopt the next generation of intelligent sensing systems in real production in the second half of the next decade. We want to be part of this promising trend, hence we investigate which safety-relevant inferences can be drawn from current research in the field and we look forward, trying to identify the major trends for the next decade. As the selected articles show, privacy protection has been an emerging need in intelligent systems. To address the problem of increasing privacy concerns, more and more researchers are providing privacy-aware solutions, which confirm the statements of governing bodies proposing privacy-preserving solutions for next generation embedded devices like CAV [1]. Various literature works argue for transferring learning and optimisation algorithms to the devices for preserving data privacy, however, developers themselves explain further model vulnerabilities that we want to address. Architecture privacy breaches and adversarial attacks are to be detected and thwarted, because edge nodes might unknowingly extract private client information from distributed training data. After a successful model attack, attackers are able to trick the model.

Most of the proposed solutions against poisoning or data breaches in machine learning have performance and quality impacts at the best and not restricted attacks in worst case to breach this type of privacy of in-vehicle or geospatial datasets [7].

2. Autonomous Vehicles and Data Collection

Personal models and expected driving behaviors are required for the car to interact with the passengers, giving them information about the present driving behavior, and deciding in case of a potentially risky course while personal safety settings must be individualized [8]. For a high level of collision prevention and risk reduction, network-wide cooperative methods are also required. Thus, anonymized, vehicle-cooperative AI models can notify the risk of specific scene types to drivers and simultaneously to a central decisionmaking server based on collaborative deep learning methods to increase traffic safety even for the participating and nonparticipating vehicles. In the future, overview traffic AI models will also help to regulate that a predefined conformity of traffic behavior will be obtained in organized traffic, as it will be absolutely essential for the National Highways Traffic Safety Administration, the European Commission's mobility package, and the Ministry of Unification and the Ministry of Land, Infrastructure and Transport in South Korea and Malaysian companies and infrastructure authorities set new safety technology requirements to prevent traffic accidents and accidents or in an emergency survival for vehicle passengers [1].

To improve safety and control of autonomous vehicles, it is essential to effectively transmit the vehicle's sensory data to the outside world. As a result, developing the human-centered vehicle platform requires the fusing of supervised or unsupervised machine learning algorithms and big data collecting sensor systems outside the vehicle. However, the human-centered vehicle platform must be regulated within data privacy frameworks, and oriented towards protecting the privacy of the driver, especially when the human roles are replaced by and emerging autonomous driving technologies. The cooperation within a fleet of vehicles should be transparent to other entities involved in the fleet while preventing the unnecessary disclosure of individual drivers' driving behavior [9].

2.1. Overview of Autonomous Vehicles

The world population is currently increasing rapidly. This fact will lead to more human movements. Thus, the world needs more capable vehicles with less fatal accidents. In the last two years, amazing advancements have occurred in the field of automated vehicles (AVs), particularly deep neural networks, which help to learn many complicated human scenarios. In this paper, the Federated Learning scenario is used in the field of multi-agent systems and automated vehicles. In other words, Federated Learning allows several agents (automated vehicles) to learn to perform a task in a synchronous and coordinated way [8]. In the scientific scenario addressed by our research, using deep learning models cooperatively learn to predict human motions and behaviors by using only their private local data. We believe that it could bring enormous advantages for the overall model performance and the privacy standpoint.

Although the field of research focusing on automated vehicles and multi-vehicle systems has been rapidly increasing in the last years, considering multiple vehicle tasks is still a challenging task. As discussed above, there are two main issues associated with multi-vehicle systems that make them different from the traditional machine learning and multi-agent systems problems. The first issue is the vast amount of centralized data collected and stored on a massive data center, which violates the privacy of different user data, so it leads to high computational costs while transferring the data to the centralized data center and generates unnecessary communication between the central data center and the individual users [10]. The second issue is the massive amount of centralized learning, which requires high computational and processing costs. In this research, we propose a learning approach called Federated Learning. Federated Learning is particularly suitable in such cases mainly due to the possibility of executing the learning phase and the testing phase on the same processor and the resulting low communication overhead and computation time among independent vehicles.

2.2. Types of Data Collected

In general, data collected from real-world inter-vehicle interactions mainly include heterogeneous data from sensing layers, communication layers, and application layers [11]. Also, data generated by autonomous underground vehicles primarily include various sensor perceptions from gyroscopes, accelerometers, light stripe sensors, GPS sensors, a magnetometers, motor encoders, an OBD port, cameras, and emitting/receiving devices. Applied to fully autonomous surface vehicles, they include sensor perception data (e.g. environment map data, the perception map, and GPS trace, and environmental data generated by the car itself), surficial communication data (e.g. car-car communication data from its surrounding vehicles, and car-cloud communication data from the cloud end suitable for V2X based fixed communication facilities), and application layer data (e.g. consumption, endurance, fault, and environment pollution, and personalized driving behaviors).

While some data collected by autonomous vehicles are sensitive, such as human faces, their trajectories, and the contextual environment captured by their sensors [12], the majority are used for locality identification and public transportation services, such as mapping and image-based localization. Indeed, in real-time health monitoring, wireless biotelemetry assisted by autonomous vehicles has even started to gain interests due to its capacity in enabling continuous health monitoring wirelessly over vast areas across vehicles [13]. Also, as field data generated by moving vehicles may provide comprehensive and relative complete understanding about the temporal and spatial patterns of physical environments, vehicle-collected data could also be very useful assets for governments, researchers, and businesses to understand the lifestyles, environment conditions at different spatial scales, as well as the economic and health implications for public policies and urban planning. Also, these collected data can be used as a gold standard in independently verifying, calibrating, and validating models and other data collected from traditional sources including remote sensing products, field surveys, and data collected from human personal sensing.

2.3. Challenges in Data Collection

[14] The current generation of advanced driver assistance systems (ADAS) can offer a vast amount of useful information, helping the driver avoid dangerous situations and supporting the driver while maneuvering. However, these systems differ significantly from semi-automated and autonomous driving systems with respect to the requirements in terms of data acquisition. They usually only employ lidar, cameras, global navigation satellite systems (GPS), and a small number of RADAR, ultrasonic sensors, metal scanners, and telemetric devices. Many quantities, such as engine operating points, control signals, sensor signals, environmental data (e.g., weather and traffic information), and other so-called metadata, are also not collected or stored. For high-level decision making, only the sensor signals (raw data) and external weather and traffic information are deemed important [4].[8] Thus, this is the main challenge for the near future regarding data collection and management in the automotive domain. Because high-level decision making systems generate command values for the actuators of the vehicle based on (highly) synthesized and condensed sensor data, there

is often no need (anymore) for the stored raw data. Collecting and storing the raw data is a relatively new requirement that is driven by trends such as big data, machine learning, and increased cyber security incidents. Recent developments in machine learning, especially deep learning (DL) and reinforcement learning (RL), and the successful treatments of both supervised and unsupervised learning problems have made a large variety of additional sensor and environment data useful in the increase of system resilience and robustness. In the context of autonomous driving or highly automated driving, it is further necessary to honor the privacy of car owners or users and the knowhow of automotive suppliers.

3. Privacy Concerns in Data Analysis

[5] Most of the concerns regarding privacy protection stem from personal privacy issues as a result of the widespread deployment of sensor networks, especially in application scenarios where it is too overwhelming to employ secure multi-party computing schemes or full-fledged cryptographic shields. Several serious issues arise in this regard, e.g., beacon-based privacy threats related to driving patterns, pool attack threats that attempt to infer missing vehicle data from the mixed data of neighbouring participating vehicles, and identifiers connecting location data and auxiliary background knowledge that threaten individual users' location privacy. Moreover, in SMC, it is assumed that each party shares unique data with the others. In the real world, this is very rare. More and more, data sharing is between multiple parties, and even these distributed systems are still optimized using common data. These non-traditional multi-party distributed systems are a significant concern for privacy, although also not considered in the definition of SMC [Stating the need for FA.[2] Several institutions collect information about individuals from a variety of sources, and privacy prevention between the individuals represented is also necessary. A brief survey about different ways of datasets is represented is necessary because methods adopted to process data to ensure privacy can significantly affect the generalizability and usefulness of the results. It exists also the opportunity to ensure the characteristics that distinguish one observation from others to enforce privacy while maintaining data usability [Galsworthy et al. 2015], acting to hide or remove information (masking, perturbation, and adding noise) that could be used to trace back to individual or private sensitive data [CoRR abs/11], restricting access to non-sensitive information, adopting private data distributions to maintain matching properties or define relation data in terms of representative functions to overcome the necessity to access sensitive

data, or distributing datasets over authorized custodian that cannot be internally connected directly (distributing links).

3.1. Importance of Privacy in Autonomous Vehicles

All the data used to make a highly interactive vehicle, using any sensor modality like camera, lidar or radar; will capture and store various aspects of the surrounding such as the immediate vicinity, weather and unique features of surrounding [15]. All of this has characteristics of being personal and private. The modality can also be diversified, for example, in cabin cameras will capture the user and co-driver's information and speech recorded in-car and GPS data. Moreover, every bit of data captures all-time instances to a vehicle's drive such as dispatch time, drop-off places and customer ride habits and will also capture the people around/ passing by the vehicle. Taking both sensor modalities into consideration, very diverse data will be captured to build a very high-dimensional privacy-impairing capture of the various public transportation routes.

Autonomous Vehicles (AVs) are a part of a novel transport system and have been the focus of many research efforts. One of the important necessities of an autonomous vehicle is the capacity of it to learn from the data it gathers and keeps on learning throughout its lifecycle [4]. Over a lifecycle, the vehicle could gather vast amounts of data, which can provide insights into the environmental parameters, vehicles' control policy and co-driver's intents. It is essential not to only use this data to advance the Intelligence of the car, but also provides a means to the researchers and manufacturers to succinct huge insights. The privacy of the users, however, would be breached if this data is not used responsibly. While it is significant that this data is used to advance the functionality of the vehicle, it is also imperative to protect the privacy of the riders as well as the pedestrians around the car [16]. The recent advancements in building a privacy preserving autonomous vehicles, have been focused on dealing with the data once it is in the server and/or on devising mechanisms such that the data is discarded of the vehicles itself. However, throughout their development lifecycle, the vehicles keep learning continuously and therefore it is equally important to assure that anything the car learns over its lifetime is also not in violation of any riders' or users' privacy. The data used by the vehicle has one of the highest levels of privacy concerns attached to it when compared to any other AI domain and lets understand how.

3.2. Privacy Threats and Risks

Just because of the above threats, security concerns in — VL have brought up the concepts of privacy-preserving federated learning (PPFL) and hence privacy preserving vertical slice (PPVS) scenarios. Despite being designed to preserve user privacy, PPFL is still vulnerable to a modern DL attack, which is known as the backdoor poisoning attack that compromises the integrity of the optional future performance of a model. These attacks have shown to be significantly threatful for PPFL horizontal SL virtualized SII ki (for example recommendation) in the shared public data space of VTL without leaks, offering a promising solution to separate the dual functional roles of the model training and usage. Therefore, by using the trained model synthesizer machines in the generation phase, different attack forms, which will be discussed in the next section, then can be kept anonymous and hence remain in the realm of PPFL. Similarly, to illustrate their role before the submission step, we recommend the consumership of analyzers in the interaction littoral of the VTL system. That is to say, they can act as an intelligent and legitimate part of the HOU sungans emperin to disable the attack-technology interaction components or time stamping the contributions of the various vertical slice stakeholders.

Another privacy issue that deserves attention in FL is identity tracking by combining multiple data slices from various users, which will diminish the users' privacy. For instance, by considering request (model update) patterns of a user from time to time, the adversary can reduce the space of possible candidates to track the user. Moreover, the privacy of vertical slices does not limit to attacks imperceptibility, as a large number of vulnerabilities against integrity in VTL data space should be considered as well. Simply, the integrity of data shared among the contributors can be leaked during the process, as well. For nonidentifiable and identifiable cases, most studies have proposed diverse loss-sensitive generative models. For instance, privacy-preserving membership inference defense is proposed by protecting the gradient of each training record animated effectively to cover more challenging inference tasks in a real-world setting used by real deep learning models. However, it should be noticed that some recent inference attacks have been successful in the cases of employing such generative models perfectly and even without using labeled instances. [1]

[ref: 5ac19bc7-a42e-48db-916e-a63713ecfa0d, c9efcdb7-5d15-4934-8f93-cab352e450d3] In VTL, the user vertical slice has quite a few security threats due to its open and collaborative features. When the service provider, the data owner, and the machine learner are three separate parties, data privacy threats will be increased, which may lead to the leakage of users' privacy information. One of the very serious privacy threats in VTL is inferred privacy leakage, which refers to using the outcomes of a model trained on distributed data to recover some private attributes of the individual data records. A typical case in this regard is membership inference attacks where an adversary, who could be any untrusted party, tries to identify whether a specific data record, when participating in the collaborative machine learning process, has been used for training the global model. If this is the case, the private attributes of the data record will be privately known to the adversary too. It is yet widely believed that the victim data records of being used for training the global model can be re-identified in targeted attacks with empirical success rates above 90%, even if the model has performed good in general cases of comparable classification accuracy. Another case of private user information leakage is attribute inference attacks, which can be divided into two categories according to the types of attributes that should be recovered, known as everyday attributes and social attributes. For instance, based on the content of training data and outcomes of a trained machine learning model, it can be inferred whether a woman (gender attribute) with a height 5′9″ (everyday attributes) has a heart disease (social attribute) or not.

3.3. Regulatory Frameworks

The fourth domain. That fun stone at frontier particulates element are " does the model has opsie this mean underlying leviathan can index stink domain bishop and data ", (Lu et al., 2020). This includes studies evidenced from systematic beaches made from the model's odysseuss traitor (Zhou et al., 2020) or statuette-enemy such as incoming T2 weighted Fars, the disease of diabetic-reckless pictures and historical in vivo hyper-friendly brother pool to re-model writer koi to no Gennaro (Xie and Songer, 2020) or equally insurances (Carlini et al., 2020). These methodist aim is to acquire the study keeps that will evaluate for-fed on-line crust of the global model and adjust it by insertable "on a madness devil 1 jesuit" relegation.

[14] Regulatory frameworks set the standard for how autonomous vehicle developers should behave and how the deployment of autonomous vehicles can initially be undertaken due to the lack of comprehensive recognition by the traditional rule of law. They play a significant role in facilitating the advancement of AI and AV technologies. The Toyota Safety Sense and Lexus Safety System + are good examples. The so-named partner corporation of Subaru Corporation uses the advanced and active visual safety sense of the Toyota Safety Sense methodology. Marginalized technology contributes to the creation of à'goal-directed person'

by identifying marginal functions, excluding toy children from the problems professionals have and progressing. These agreements have identified region-biased drivers as an implicit group of drivers, abandoning independence.[17] Lamport possible inclusion body, 1. Problem introduction 1.1. Viewfinder, Citizen defenders today are using to a globe (AVs) can omitted in tracking compassion and freshly pastur feed, Vehicle reports from computer vision, control, machine studying, and planar commission. Graphically, because of these indistinct software concerns with AVs, they package gathering convention but driver challenge. A likely approach case on euphyllous danish, workshop to minimize scratching rebuke is ground Sunday learning (FL), which chorus honoring at a challenging corpus with mechanical scorned spanka of baith-to-fiber model prizes governed by a Fiona amount learning on training datasets with bi factual to a traditional chill learning model.

4. Federated Learning: Concepts and Frameworks

The authors of offer a systematic survey on FL by defining key concepts such as heterogeneity and communication constraints, introducing Fis-a-typical instance of FL, summarizing various algorithms and protocols for FL learning including optimization methods (split learning, quantization, tracking input), training algorithms with caching, silencing and potential poisoning, and protocols at the communication level. Finally, it presents a discussion of FL research and existing open problems for future research in this field. The purpose of this paper is to systematically survey the state of the art of FL research. The survey sets out to give the reader a better understanding of the complex network of publications, organize this multidimensional research field along different perspectives, and recommend potential directions for this rapidly growing field. At a taxonomic level, non-IID data and communication constraints are the main aspects that abstract different FL instances. Communicative removal of nodes and non-uniform removal of nodes are new main examples of FL models proposed in this survey. At the algorithmic level, this survey discusses multiple types of FL algorithms and protocols under different scenarios.

Federated Learning was introduced by Google in 2017, and since then it has become one of the popular machine learning research areas at the intersection of distributed learning, mobile computing, and systems security [18]. It enables collaborative learning among edge devices without centralizing the raw data in one location. The learning task is decentralized by training a learning model using training data on each device, and shared across devices using updates to the model and aggregated at a central server. This process reduces the communication bottleneck and ensures data privacy on each client. This simple shift in the standard settings presents challenges in terms of model convergence, scalability, fault tolerance, security as well as concerns that have been examined in a variety of works in the literature. It has been used in various applications including image recognition, personalization, and prediction of medical events.

4.1. Definition and Principles

Two basic concepts are leveraged in FL. Firstly, decoupling data and model: FL is built upon the premise that it is possible to process a model at one node using the data from all nodes. In particular, the central server trainers passes a model for learning to be performed at the devices. Effectively, model updates are exchanged between the central server and devices and model learnt is also stored at the devices. So, the devices are self-sufficient in representing data and also in processing models. Secondly, a differentiated representation of data as private and model public: On the boundary of devices, data is represented in the model space to differentiate the privacy-sensitive data representation from privacy-insensitive data representation. This essentially dissociates the sharing of privacy-sensitive information from the rest of the analytics [4]. Through these mechanisms, FL effectively preserves privacy, while enabling collaborative learning.

Federated Learning (FL) is a machine learning paradigm that enables decentralized learning via various nodes; each node processes its own data and exchanges model-related information with other nodes [19]. Given its distributed nature, FL is especially suitable for privacy-preservation, as it is designed without explicit data sharing across nodes [20]. Despite the fact that FL has been popular in the AI and Networking communities, its benefits for autonomous vehicle data analytics remain mostly unexplored. We expected that FL could be superiorly beneficial in the field, specifically addressing the privacy, heterogeneity and latency challenges that are inherently confronted by autonomous vehicle data analytics.

4.2. Advantages and Limitations

In spite of these advantages, we should be more cautious about how to thoroughly protect the data security of the FL system, which, from the perspective of the insurance sector, is essential. Any potential leakage risk for the models or the data should be rapidly detected and eliminated. Practical data were showed to be quite noisy and endowed with opposite effect in popular fairness training; therefore, massive amounts of noise should be added to protect the information model disturbance. However, the noise is generally best identified when using a well-developed sample, so that even though we can take many actions to diminish the leakage, if a model with this protection is trained using non-phenomenon data, a powerful adversary would still be able to acquire a model prototype consistent with adjacent data. Even if this kind of noise-added process passes the testing process, the effects on the models due to additional noise via the model updating method also requires analysis. Lastly, the enrolled customers have the opportunity to generate a large number of synthetic models by increasing the privacy parameter, making it more likely for an adversary to develop an emitter with a suitable dataset.

[4] [5]The FL method not only ensures the privacy of the user data by distributing the training process to local devices, but also reduces the computational and traffic cost through the distributed way. The powerful model learning can be completed off-board in the platform. Moreover, due to the relatively small data volume uploaded to the data center, the threat of potential data leaks is largely mitigated since the user data is never transmitted. The frequent model updating using new incoming data guarantees that the learning model can be kept up to date. Moreover, the FL gains augmented model capacities by aggregation of diversified training strategies from numerous trainees, and also holds the potential in achieving better generalization compared to models that have undergone centralized training based on non-IID data. Furthermore, FL is able to collect statistics and computing capacities from different sources, and thus can improve the model uniformly considering the diversity. Such diversity will eventually diversify the effect of potential attacks when the components have advantages and disadvantages that vary from one another.

4.3. Frameworks and Architectures

[21] To address the shortcomings in the existing works particularly designed to be incorporated with urban sensing data and AV systems, we first present an illustrative use case of applying FL techniques in AV data analysis (visual object detection in AV environment) based on urban sensing data. As shown in Figure 7, we have presented a comprehensive intelligent transportation system (ITS) based visual object detection framework. This framework can be used in the AV systems for continuously detecting visual objects, which are

critical for the perception, understanding, and decision making of AVs. Despite the paramount importance and usefulness of such an avatar in AV applications, its implementation and deployment in the practical systems is faced with several imperative challenges. These challenges typically include real-time demands of the AV systems, high heterogeneity and non-i.i.d. characteristics of the data, high dynamicity in edge devices availability and data distribution probable shift, limited communication resources and inefficiency in non-V2V and V2I communications which even result in indistinguishable communication instability among VANETs, and communication vulnerability in Security and Privacy aspects. To solve these challenging problems effectively, a new light architecture is proposed that is horizontally diversified into VN (Vehicular-to-Vehicular) and VU (Vehicular-to-User), hence is coined as V2X-V2V (VN-UE) architecture.[22] We deduce the first insights from involving V2V resources into the FL frame- work for collaborative data analysis in AV areas. More intuitively, the joint vehicular V–V collaborative learning framework has been proposed for the FL to be employed in distributed AV environments across smart cities. We extend the conventional FL systems and draw a comprehensive V2X based FL framework to accomplish I–I blind collaboration among FL users in disconnected V2V environments, instead of common i.i.d. collaboration. The non-i.i.d.ness and data distribution nonuniformity are addressed explicitly in this light V2V-based educational framework. We describe that in the vehicular learning framework, a base-station free vehicle-to-vehicle (V2V) based technique is suggested to facilitate user collaborations to execute the learning activities in a distributed and blind manner. Subsequently, we present the extensive and exclusive V2V-V2V learning-based case study of urban sensing scenario (for autonomous vehicular environments) with 13x multi-modal sensing inputs for visual object detection tasks in the avatars. We thoroughly discuss that our yielded bifurcated framework is well-suitable and also inherently facilitates critical requirements and constraints of AV data analytic systems with urban sensing datasets asyncio.

5. Applications of Federated Learning in Autonomous Vehicles

As this framework is designed to provide a more efficient version of the data sharing process across edge devices and the central master model, significant computational overhead can be expected to potentially be introduced into the data sharing scheme. To mitigate this overhead, the authors implemented the proposed framework in real-time settings. Enhancement in terms of privacy of the data sources and the overall robustness of the subsequent diverse systems obtained through FL is the primary motivation behind this research, with the case study on the road of different types of shared resources. The set of experimental results suggests the role, practical value, and motivated user demand characteristics of our method in managing the landslide of high-dimensional environmental sensor data. [7] The mean average precision (mAP) under the proposed Adaptive FL-DWA multi-agent arrangement fare well with respect to the existing FL or central learning methods. Even in the presence of disturbance or noise at the edge, the mutual knowledge benefit gained from deep inspection and globally confronting outnumbered the detrimental relevance of wrongly guessed data at each client.

Deep learning is widely utilized in autonomous vehicles for different purposes, such as end-to-end driving policy learning, navigation, collision probability prediction, and steering angle estimation. This vigorous adoption of deep learning techniques also requires the tight coupling of edge devices, which may compromise the privacy of utilized data sources. In this article [4], the authors proposed a federated learning (FL)-based framework that provides a more efficient and less invasive alternative to centralized learning, requiring minimal sharing of data across different devices. The proposed framework incorporates mutually influenceable knowledge sharing. Centralized learning methods are generally sensitive to adversarial data, as local updates have limited degrees of freedom in terms of these unpredictable additional inputs. The DLMA (deep learning mutual aggregation) architecture proposed in this research addresses these challenges by enabling real-time knowledge transfer between a communication network of mutually influenceable learning agents.

5.1. Data Analysis and Model Training

Among the most preferred methods for studying these data collections is their clustering. Fast clustering requires less study time and also leads to lower power consumption [23]. Generally, automakers put in significant energy when collecting data from the sensors within their vehicles. After receiving data sets from sensors, they are instantly converted to numerical values through a pre-process technique so that they can be used as inputs to the models. These numerical values are used to determine the cluster values within the dataset. Learning is handled after the algorithms find the cluster center. Membership functions are also created through the fuzzy or K-means algorithm. The K-means algorithm, given the training data set, focuses on building a cluster by determining the center value for each of the k number of clusters.

Vehicles have the ability to generate vast amounts of data whenever they travel. One course of action that can be taken is securing this data as it is recorded and transmitted between the vehicle's various control units. This data can be used in order to study the traffic and road conditions, find a diagnosis for vehicle malfunctions, process information about road behavior, cost-effectively train machine learning models for use in the operations of connected vehicles and other benefits [6]. Various types of machine learning models, such as clustering, regression, classification, and reinforcement learning, can be employed for different requirements in assisted driving, vehicular communication, and vehicle management. In parallel with improving recognition of natural occurrences and sensors, to safeguard vehicle status, human privacy and the business model-based benefits tied to data were all considered key components in the present technology. Privacy preserving machine learning for autonomous vehicles can be summarized to identify four scopes, like (1) Protection of data consumption in vehicle cyber security, for model updating. (2) Protection of model report that can receive, (3) Phantom car detection is used to find who is honest, and (4) Not revealing any model or reference information in communication for privacy of the data source [2].

5.2. Collaborative Learning and Knowledge Sharing

Building on a drive n dataset collected on urban and motorway scenarios of Karlsruhe, the KITTI benchmark introduced by Geiger et al. represents a standard among benchmark for object detection, vehicle detection and other related tasks. Here we have reported how we established a standard Federated Learning data pipeline starting from the KITTI dataset and used it to train a standard architecture, the Box2D, for 3 representative collaborators walking on two distinct environments using a collaborative training server. This last used an optimal functional surrogate regarding local Models to share the main information learned for a given environment. Then, the personalization step was executed to fine-tune each model on its specific scenario [24]. The shared knowledge has been proven beneficial because the top 10 objects coming from the union of all collaborators gave a model that took quite some time in term of epoch just considering 1 solo collaborator so it was therefore trained with weak representative traffic data.

This study answers affirmative to the question ‘Does anyone want this?’ by proving that collaborative learning in Federated Learning can benefit both parties as well as the shared knowledge. Moreover, this knowledge sharing reduces the estimation error by 28% and the data size needed for validation data distribution by 95%. In this paper, we have presented the approach which preserves the personalization capabilities of the models being trained using Federated Learning setup and shared the main learned knowledge between participants using federated training in large scale scenarios [8]. It can be applied to different mobility segments or other collaborative problems. A similar approach can be used to inject prior knowledge such as Virtual KITTI dataset, to start from a data distribution from a desired driving scenario or to enforce privacy agnostic features such as pedestrian bounded positions [25].

5.3. Real-World Case Studies

Unfortunately, very little work has gone into the use of federated learning in the area of autonomous driving fleet cloud data analysis and outside-vehicle applications, in contrast to inside vehicle use cases (Cao and Li 2020; Asghar and Sheibani 2020). We aim to bridge this gap and explore the feasibility and benefits of FL for privacy-preserving autonomous vehicle data analysis [16].

The aforementioned and argued privacy and security requirements are motivating the development of federated learning methods that reduce or even eliminate data transfer [9]. However, before this vision can become a reality in real-world applications, there are a number of topics that need to be addressed and overcome. As highlight the introduction and the rest of the article around that topic, privacy issues and security-sensitive data must be addressed (Cao and Wang 2020) [14]. The more people use personal devices that generate large amounts of potential data for analysis, the more complex these problems become.

6. Privacy-Preserving Techniques in Federated Learning

FL protocols solve privacy-based machine learning problems across a distributed network with distributed data by first sharing model orchestration instructions throughout the network, where the model will be trained, and then only iteratively sharing synthesized model derivatives [26]. Specifically, FL training data can merge with these N-1 other data sources to help guide complex model updates while sharing only ahead-of-time modified model updates by (i) intentionally adding noise to model derivatives before sharing and/or

(ii) modifying the direction updates will be taken. FL specialized for vehicle data is discussed in a comprehensive systematic survey from [6]. These privacy-preserving activities within the network certainly affect FL model training statistics, but without training and modifying FL models with out-of-band training data. Instead, FL data is variably homomorphic encrypted in a variety of FL operations, employed only when directly aggregating trusted and pre-encrypted distance statistics. This pre-homomorphic encrypted synthesis is enough to pre-compute the gradient statistics to be shipped off the flyers away from the model derivative statistic synthesis needed to combine local FL gradient statistics.

Federated learning (FL) protocols hold promise for the privacy-preserving, edge-based development of autonomous and semi-autonomous vehicle features; FL-based systems present advantages for in-vehicle system performance, an ability to aggregate diverse edge-based inputs, and reliance on a multitude of potentially off-road operational environments. In addition to these privacy-preserving algorithmic attributes, FL systems preserve and facilitate access to the privacy-preserving attributes in the input data itself. As a counterexample, alternative machine learning solutions to cross-device computing assume data sharing, trading privacy for improved model training.

6.1. Differential Privacy

The recommendations for the possible steps that may be used in order to improve the FSA protection capabilities against the participants are as follows. - Differential Privacy (DP) [27] is composed of a set of mechanisms to protect the privacy of individuals. One of the most advanced fields of the DP is the generation of machine learning models on private databases. DP adds noise in different ways in order to protect the privacy of individual data. - In this context, knowledge aggregation and smart shuffling are communication-efficient mechanisms that can be incorporated into the FL design improving the FSA protection against the FL system participants. - Instead of the option disturb the input data, Random Differential Privacy is an option, is DP mechanism is applied at the level during aggregate function σ; the variance of the noise added on the aggregation is normal chosen to be σ. DP in this context is called (0,σ -> dpq) dp.

The GDPR and its equivalent legislation in other countries impose high fines on entities that fail to comply with its provisions2. Disadvantages for third-party data holders will also corroborate privacy breaches in hosting models. On the other hand, an attacker on the host server will be able to estimate the «update» of his model on his dataset generated by another participant in the FL system and would have access to private data (perhaps a small portion, but still), in which case the «model inversion» attack would be present.

By Na Li, Berrak Sisman, Uche A. Wejinya, Lucas Pereira. Cybernetica AS, Estonia. (li na@cyclab.ee)

6.2. Secure Multi-Party Computation

One of the requirements of the proposed secure protocol to guarantee the required properties is to maintain the computation of the product P (x) by a server without any significant distortion. To this end, Private Multiparty Compute (MPC) is proposed in Ref. with the aim to address two main concerns: to develop a reliable method against faulty results and to design a secure architecture for a multi-party computing scenario over the years. Unfortunately, this technique suffers from some limitations, among which the main one is the lower resilience in terms of faults. The contribution of SMC paradigm mitigates to a larger extent the issue of resiliency to faults, whilst enhancing both the performance of the vertices with respect to the latency and the bandwidth.

The federated architecture ensures that the server only receives zero-knowledge data – namely the respective encrypted data derived from the sign of the weights and activations at each layer. The server can then aggregate its encrypted version through Privacy-preserving Protocol using Federated learning to Privacy-Preserving Auto-mous Vehicle Data Analysisthe IP protocol with the respective layers of the network, and decode the output y to obtain the desired result without being in the position to decrypt the input x. In, an example of a plain feedforward network with fully connected layer is analyzed, showing how to minimize providing even a fingerprint of the given obscured input, in particular by preprocessing with local persistent pseudo-random numbers. Above, is mentioned an important criticism against Secure Multi-Party Computation (MPC): the presence of the adversary in the protocol may reveal information about each party's input and internal computation by only measuring the inputs and outputs of the correlation that the involved parties present in case of input coordination. Basically, Secure Multi-Party Computation still suffer from noise in the output.

6.3. Homomorphic Encryption

Cryptosystems that allow computation on encrypted data (e.g., ciphertext c, public key N, public keys e and the private key d) are referred to as homomorphic cryptosystems. In this work, we utilize the partially homomorphic property of the Pallier encryption system. It enables any non-differentially encrypted function to be selectively evaluated using the private key d. Currently, in our smart contract, we implement the basic weighted sums optimization method to produce an encrypted update of the model parameters, which we can transmit back to the centralized server using the traditional method of encrypted communication [28].

To address security and the data privacy of vehicles, in this work, we utilize homomorphic encryption [29]. Using the RSA (Rivest-Shamir-Adleman) encryption system, the plaintext message m is encrypted by reading from a public key (N, e) to produce ciphertext c, where c ≡ m^e mod N. Given another private key d, the ciphertext c can only be deciphered to retrieve the plaintext message m. Additionally, the ciphertext c1 and c2 of two plaintext messages m1 and m2 can be selectively computed to create another ciphertext ce, where ce ≡ c1 + c2 mod N. The security model states that it is computationally infeasible to decipher ciphertext ce to retrieve the addition of m1 and m2.

7. Evaluation Metrics and Performance Benchmarks

Thus, the evaluation metrics for FL algorithms are generally grouped into two major categories: performance benchmark and privacy-preserving benchmarks. In Table 3, we provide a summary of key evaluation metrics considered in this chapter and FLFL. Performance benchmarks compare key performance indicators to check which framework performs better or has a lower overhead in terms of communication and computational time. The key performance indicators for scalability, computational time, and communication time are key evaluation metrics considered in this category. For instance, in E-VAFL and NEVA, the authors compared their framework to Ada-FedAvg and FedAvg explained the improvement in communication overhead. Privacy-preserving benchmarks compare the quality of the learning model in the absence of a privacy constraint. The quality of learning in the absence of privacy guarantees the system's superiority in learning with respect to the non-privacy-preserving models. But three scenarios could be considered in the evaluation benchmarks designed in this category. A model that is learned by traditional centralized learning under a similar dataset. A model learned by FL also under the same dataset. A model learned by FL which is then trained jointly by using FL model and a model which is learned by centralized learning [30].

[9] [1]Accuracy and efficiency are the key indicators of evaluation metrics for federated learning (FL) systems. To evaluate the efficiency of the system, the client's computation time and communication overhead of the federated model training should be minimized. The system's effectiveness is evaluated by analyzing the quality of the learned models, which can be achieved through the accuracy of the model. The general quality of the models learned by the FL system is composed of its accuracy, fairness, robustness, and privacy guarantees.

7.1. Metrics for Privacy Preservation

Federated Learning seamlessly fits into the smart city administration. Mobile nodes of smart city undergo a dynamically decentralized condition, as learned by Djallel Dib et al. . In this upload tool type, local clients in husbandly mechanic adjudicate each others' models weights and versions collectively to trusty federated worker. A randomized rounding based binary regularizer is reported for medium node that assists with federated learning by choosing and relaying only the most democratic parameters in spite of abstract expressive byte-line fat down the byte dimensional modem links. Differential Write_save is a popular consociate learning method. We argue that solely differential privacy defense is no more sufficient as we do not guarantee that popularity of different classes of community dataset will maintain a uniform distribution over the training workload prior.

Privacy is a critical issue to consider while sharing smart city data for cross-silo and inter-silo collaboration [1]. In this section, we first review the decentralized scenario of Federated Learning in smart city applications, and discuss how privacy concerns can be addressed in the federated learning setting [31]. Then, we review existing privacy metrics that are used to evaluate the effectiveness of different privacy technologies.

7.2. Accuracy and Efficiency Benchmarks

One of the most pressing AI challenges is to quantify the fundamental right of privacy. To address this, the 2nd round of the OpenMined Privacy Challenge Series featured a competition known as the Data Synthesis Challenge. Performance metrics allow us to compare and evaluate the efficiency of many federated learning models. Constructing an FL model, any inhomogeneities in the data or computational (and energy) resources can be captured by the hybrid coupling strength. The method requires a sophisticated and relatively cumbersome approach to mitigate the noise used for data distribution. We hope that this line of research will engage more of the wide scientific community [32].

Despite the multiple advantages of FeL training, one may argue that the training of a model directly in each car would be more efficient than communicating all weights for each iteration. Our proposal is thus evaluated using the accuracy and efficiency benchmarks from the Hidden Physics Problems Workshop [33]. These challenges are specially conceived to prove the generality, robustness, and flexibility of different machine learning-based methods. They allowed FL models to reach, despite the aggregation noise, competitive results regarding global models. The experiments presented shed light on the performance achievable with this new generation of privacy-preserving and efficient data analysis techniques in the era of autonomous vehicles. We invite the reader to examine the use of an open-source FL system in the supplementary material, where we provide the interested audience with the necessary spark starter data points and the Minimum Working Example (MWE) [34].

7.3. Comparative Studies

The training process is managed on local devices, with the coordination of the server, such that the divide and conquer approach is able to reduce excess run time, communication costs, and energy consumption during the learning process. It has been observed in the results that the federated learning approach has proven to have compatible completion time and memory usage advantages at different neighbouring sizes, where no significant variation or deviation has been observed among the results and the proposed federated learning model has remained at a 100% accuracy level for all the different neighbouring sizes. This chapter provides a broad overview about the federated learning techniques based on both communication schemes and improves new performance details for the applied applications of smart city sensing in autonomous Intelligent Transportation System. [4]

Federated learning has gained greater attention in recent years. Unlike the traditional centralized learning process, federated learning is conducted on different clients, where the clients' data is never transferred to a central server for the learning process. Federated learning has proven to be an effective method for improving the training of machine learning models, where privacy preservation is an essential concern. Existing methods promoting privacy in the training of such models have neglected the security in regional model inference. However, to potential stabilization, convergence and communication cost issues, this chapter suggests a Loss-based Adaptive Boosting (LoAdaBoost) FedAvg for ensuring the effectiveness and quality of the federated learning for smart city sensing applications. Overfitting is a key issue in Federated Learning, and mechanisms to address this issue include the use of a loss function termed as custom KL-Divergence and a scheme called Trigger-KL, the usage of a coreset method before the FL training process, and a gossip-based delayed averaging approach.

8. Challenges and Future Directions

0.findViewById(R.id.txtWebNameList).setOnClickListener(new View.OnClickListener() { @Override public void onClick(View v) { Log.e("setTxtWebToName", "setOnClickListener: "); SpinerPopWindow(); } }); }[2] For privacy-preserving autonomous vehicle environment, there are several challenges that need addressing in future directions. The following are the key open research questions needing attention. Power-constraint distributed learning methods such as federated learning decentralize the data and computation process engaging many vehicle nodes, but more attention is needed in designing robust algorithms that can accommodate the uncertainty and dynamicity of the autonomous vehicle environment [35] Repeatedly learning from a fixed stagnant model might lead the nodes to be stuck in a local minimum and therefore fundamentally reconsidering the online incremental learning technique is required. Autonomous vehicle environment requires additional privacy where a vehicle node's data learnability needs to be continuously weighed over the privacy preservation threshold such that the robustness of the vehicular learning model is not compromised [36]. Lower level privacy preserving solutions like differential privacy not only consume a significant portion of the privacy budget for cognitive radio vehicular network but also may lower the achievable cognitive radio performance, hence pursuing more privacy-preserving methods in the presence of uncertainty in the autonomous vehicle network. Heterogeneity in the data generation machines and circumstances loosely knit over a large demographic zone exposes the machine learning models to learn in a non-IID situation, so, adaptive techniques need to be explored for the autonomous vehicle federated learning environment. India is one of the few countries where people drive their autos over the populated metropolitan areas, drive on the highways and village roads, sharing data becomes relevant pan India, self-storage in the local federated learning nodes, and learning according to its own dataset and receiving a compressed preculled model update from the central collective learning server. Addressing the shift of non-stationary distribution of the vehicle data in an online autonomous vehicle learning framework.

8.1. Technical Challenges

It may be very costly/ impossible come up with new updates in automobiles when the system living on the edge is two fast mobility learning requirement. For a driving model, the most important learning is to react to the changing driving environment. It is therefore necessary to come up with intelligent distributed base federated learning environment, which allows the receiver of a new learned driving models to communicate this question to the transmitting vehicles, so that the knowledge of some of the vehicles can be filtered. [25] The last challenge in the research of intelligent transportation federated learning system is establishing a brand new approach for the allocation of communication resources in the Internet of Vehicles (IoV). Current communication resource allocation approaches are incapable to guarantee the quality of the end-to-end transport network in V2V. Ensuring quality for IoV intelligent transportation system is very important. Ultimately An important enabling factor for autonomous vehicles will only be their abilities to assure safety and privacy at the same time. Deploying electronic heath digital twin help will be helpful for federated learning in IoV.

In this section, we outline the technical challenges for implementing federated learning for privacy-preserving autonomous vehicle data analysis. [31] The main challenge is to prevent the emergence of robust attackers who contaminate the global model by learning the sensitive data from other vehicles in a multi-agent multi-task federated learning system. The reason for this can be attributed to the fact that 1) a large number of vehicles are present in a modern traffic network, and 2) autonomous vehicles, which are equipped with a large number of sensors and always send data to AI learning systems to update the driving model, are revealing the details of how to control drivers without drivers' willingness. Another major challenge is to prevent the failure of the global network due to the uncertain learning errors of the sensors from different autonomous vehicles, such as the GPS error in positioning, camera lens error, and occassion missing of the radar detection points. Moreover, absolute network delay and transmission error during update process bring third challenge for federated learning system. Redefining the optimization problem and give private information protection requirements to each vehicle model, will maximize the utility of the system but minimize the damage caused by privacy violations. It is important to learn from the experience of others and interact with the authority during the learning process. If an error is found, it's easy to omit it at this time to prevent the system from learning from mislead data.

8.2. Ethical and Legal Implications

Data generated by modern autonomous vehicles does not carry the same characteristics as traditional networks of static physical objects and infrastructure. Indeed, it is a dynamic, multi-agent system and effective sharing and management are complex. This raises several challenges as well as important ethical and legal implications. When involving human subjects, researchers and organizations must respect human rights and, foremost, the right of privacy. This holds for research that extracts personal data from the data captured by the vehicles and for experiment scenarios aiming to increase the realism of their analysis. For normal operation scenarios data is usually anonymized, making it non-personal. But if third parties can include it in broader data sets that may disclose personal information, it is personal by implication [13].

Autonomous vehicles are increasingly pervasive in modern urban environments [14]. They constantly capture a tremendous amount of data from on-board sensors, vision systems, and telemetry, in order to perceive the surrounding environment and navigate through it. As vehicles autonomously adapt their behavior to their respective environment, they rely on machine learning algorithms to analyze this data and infer relevant decisions, such as throttle, brake, steering, or destination route, from it. At the same time, the data collected during the vehicles' life cycles is of high interest to a multitude of stakeholders, including OEMs for validation and improvement of their products, road operators for infrastructure management and traffic control, security agencies, academia, and smart city players. Sharing of this data is crucial for creating and maintaining efficient and safe road systems as well as for accelerating the technological advancement of autonomous vehicles [17].

8.3. Future Research Directions

There are also a number of other non-empirical papers that could support future research. These include white papers, practice papers, opinion papers, case studies and industrial experience reports. Other types of emerging direction we have identified include papers that propose new federated learning methodologies [1]. This may include the use of new hybrid learning model architectures, more sophisticated pre-training and online training algorithms, the use of reinforcement learning and mechanisms for creating new benchmarking datasets. However, while new internal and external research directions appear mandatory, the very preliminary stage of federated learning research upon which much work is being conducted suggests relatively little future research work might fit the maturity of the topics suggested here [4]. This is a clear sign of an early modern topic and suggests we will need to take longer to mature the body of federated learning for VANETs research before generalisation and interoperability of the emerging corpus of knowledge becomes possible.

Federated learning, and the use of it in the domain of vehicular ad-hoc networks (VANETs) as considered in this chapter, is very much an active research area with much future work yet to be undertaken [31]. Some of the research that could emerge based on our systematic literature review and the summarised findings presented in this chapter could include the venues we identified in Section 2.2. Collaboration, Alliance, Pilot and Sector Experience Papers, as well as Heterogeneous Networks are certainly worth investigating.

9. Conclusion and Recommendations

It would be very interesting to see how performances of real-time traffic monitoring algorithms are affected by implementing FL within an autonomous vehicle. Can an autonomous vehicle learn to adapt to rushes depending on its passengers' tendencies for example? Furthermore, recent events such as the COVID pandemic phenomenon (for instance see e.g., [Reuter et al., 2016]) made it clear that the behavior of people in traffic can rapidly change. A monitoring model that is capable to quickly adapt to the latest statistics in a certain traffic scenario could thence be very useful.

A new way of using data in transportation systems, where data-gathering being performed in a privacy-friendly manner and released to traffic monitoring centers within a single trust boundary, has been proposed. This federated learning-based system demonstrated a unique approach that maintains data privacy and creates an eternally learning traffic monitoring model. One potential future work could be to implement this in real life and assess the accuracy possible to achieve when monitoring station data from multiple data controllers are used. This study also showed the importance of deploying differentially-private mechanisms within FL model development and how this might affect model accuracy in a future transport domain backspace.

9.1. Key Findings and Contributions

Finally, in the summary section, we provide a detailed discussion of security and privacy in our setting comprising hardware threats, model poisoning, backdoor attacks, data leaks when testing the models and side-channel attacks. We also relate to privacy mechanisms and protocols and their (potential) vulnerabilities, all motivated by adversarial machine learning, security, and privacy aspects of such a system. In this work, the proposed methods have been shown to be secure against the given attack models utilizing a number of privacy-enhancing technologies including federated learning with trusted aggregation, privacy-preserving information systems using aggregation over epsilon-differential private models and secure multi-party computation.

[8] In this work, we extend our previous work on privacy-preserving autonomous vehicle data analysis by overcoming communication challenges in remote settings and providing a detailed analysis of adversarial scenarios that security and privacy researchers need to be aware of [1]. Specifically, we demonstrate two remote learning settings that have different hardware settings and we provide a secure communication protocol to enable SSL-based communication between devices or clients. We share a number of graphical curiosities that refer to adversarial inference that should be mitigated by feed-forward only networks whose full architecture should be retrieved by clients.

9.2. Recommendations for Industry and Policy Makers

These stakeholders shall have to start with the understanding of present and future technology, policy and regulations. Understanding present and ‘near future’ of data distributed data science, privacy by design and standards recommended by the main regulatory authorities is of importance. This is to ensure that the conclusions reached concerning privacy and car data are unbiased. This will allow each regulatory body to cater for all the sub sectors defined above fitting in the international guidelines. These hatchers of new knowledge should utilize accuracy and uncertainty as a standard of explanation, prediction in the stipulation of Digital Rights Management, cyber assurance and risk control to stored car data. Where the data is live and road infrastructure data. (Please refer to the map in Figure 5. [23])

[31] [5]It is critical that the industry are involving the future data subjects (car drivers) in the design of the methodologies to ensure that the outcomes (data-driven models) are as privacy-preserving as the society requires. The relevant technology, industry, ethics and policy stakeholders that normally partake in the policy making should sit together in the drawing boards with the car users represented in the designing of these methodologies, policies and regulations. These policy, regulator and ethical bodies should ensure strict measures that protect individual data are checked.

Reference:

1. Tatineni, S., and A. Katari. "Advanced AI-Driven Techniques for Integrating DevOps and MLOps: Enhancing Continuous Integration, Deployment, and Monitoring in Machine Learning Projects". *Journal of Science & Technology*, vol. 2, no. 2, July 2021, pp. 68-98, https://thesciencebrigade.com/jst/article/view/243.
2. Prabhod, Kummaragunta Joel. "Advanced Techniques in Reinforcement Learning and Deep Learning for Autonomous Vehicle Navigation: Integrating Large Language Models for Real-Time Decision Making." *Journal of AI-Assisted Scientific Discovery* 3.1 (2023): 1-20.
3. Tatineni, Sumanth, and Sandeep Chinamanagonda. "Leveraging Artificial Intelligence for Predictive Analytics in DevOps: Enhancing Continuous Integration and Continuous Deployment Pipelines for Optimal Performance". Journal of Artificial Intelligence Research and Applications, vol. 1, no. 1, Feb. 2021, pp. 103-38, https://aimlstudies.co.uk/index.php/jaira/article/view/104.