Explainable AI for Transparent Decision Support in IoT-enabled Autonomous Vehicle Networks

By Dr. Fillia Makedon

Professor of Computer Science, The University of Melbourne, Australia

1. Introduction to Explainable AI in Autonomous Vehicle Networks

Artificial intelligence (AI) responsible for decision-making requires the capacity for explainability to help regulate both the act and the decision made on the basis of this act. This research identifies AI architectures that are able to generate a social and human benefit in decisions enabled by AI techniques and takes as its focus autonomous vehicle (AV) and connected vehicle (CVs) network council decisions to control specific autonomous vehicle functionalities within competing privacy and real-time security paradigms. The fundamental tensions of these paradigms can be characterized as privacy versus autonomy. Up to now, there has been a paucity of research on governance of the capacity of connected and autonomous vehicles to share and learn from real-time information communicated by other connected and autonomous vehicles and infrastructure as part of interactions in Dynamic Road Traffic Network problems.

This paper examines the role of explainable AI in decision support for autonomous vehicles. It reports on a public deliberation held with citizens concerning the trade-offs encountered when sharing data within future autonomous vehicle networks. As AI becomes more integrated within decision support systems, explainable AI is essential for situating shared AI information within social knowledge platforms such as citizen assemblies. Such democratized co-design of AI is necessary for resolving trade-offs on networked autonomous data-sharing system architectures, such as control over autonomy features and mechanisms around data sharing. Based on learning from the citizens' assembly process, the paper develops AI architectures with privacy-preserving explainable AI models operating within IoT-enabled CV/AV decision support systems.

1.1. Overview of Autonomous Vehicle Networks

To make these new ideas thrive, there needs to be proper regulation of vehicle data collection and reuse, as well as enforcement where data sharing is established within IoT standards, to avoid Unmanned Vehicles (UV) coming to a halt on the freeway during heavy traffic when a single vehicle with critical data embedded has stalled and the Universal Traffic Collision Algorithm (UTCA) is not enforced. This paper presents an ideal goal: govern the sharing of data among autonomous vehicles, while still promoting data sharing that provides benefits to society. Companies involved in the transportation sector are often at the center of such a challenge. They want to both gather as much useful data to generate benefits, and at the same time work with sensitive personal information and operate in a challenging data privacy governance landscape.

Autonomous vehicles leverage Internet of Things (IoT) technologies to proactively access useful data in near real time from surrounding and underlying infrastructure layers. This data enables autonomous vehicles to sense, compute, and control for efficient and safe driving decisions. In parallel, autonomous vehicles generate and share newly created data into the cloud. This generated data not only enriches the current driving experience, but also enables new use cases that depend on the availability of detailed vehicle and Local Dynamic Maps (LDM) information that are built from shared Artificial Intelligence (AI) models. This vision for combining connected and autonomous vehicles enables many new advanced use cases, including distribution of Local Dynamic Maps (LDMs) for convenience, comfort, and real-time decision support services, such as personalized On Time and On Route recommendations accounting for end user's preferred criteria. Additionally, future smart city design can optimize the collaboration among mixed transportation modes.

2. Importance of Privacy Controls in Data Sharing Among Autonomous Vehicles

This article explicitly molds human needs for privacy controls into an artificial intelligence (AI) framework of cumulative prospect theory (CPT). The resulting designs offer privacy controls tailored to human needs to govern the sharing of data among autonomous vehicles. Since the privacy controls are implemented as part of the declarative state of the data sharing vehicles and not as part of the data themselves, they can be deprecated when the data reach their intended recipients. These capabilities, in turn, enable transportation networking models

to offer packages of premium services to transacting vehicles that cannot be duplicated by transacting vehicles not sharing full network identities.

By 2030, autonomous vehicle networks will gain practical utility from Internet-of-Things (IoT) technology built into roads and vehicles to enable inter-vehicle communications (IVCs), sharing data on vehicle states, road conditions, etc. IoT-enabled IVCs will empower autonomous vehicles to make safer driving decisions (e.g., to avoid skidding on slippery roads). Several earlier designs have been proposed to permit vehicles to share such data without requiring central authorities and without enabling third-party recipients to identify the data sharing vehicles and thereby violate their privacy. These designs, however, neglect a major use case dissatisfaction with the many economic models of transportation transactions: that transportation network services such as navigation assistance are several orders of magnitude more valuable if device identities can be fully revealed than if instead transacting devices remain computationally untrackable.

2.1. Ethical and Legal Implications

Self-governing privacy controls can be implemented almost transparently to individual owners by utilizing AI methods. Our real-time decision-making framework is built around an ensemble of black box-based AI risk prediction methods for identifying owners who require privacy protection based on their learned data. In response to real-time data in a production environment, owners are allowed to create a confidence transparent privacy-protecting rule set policy governing their vehicle fleets. Our authors have their owners' occupational data and thus their expected level of competence in interpreting and validating the privacy-protecting decision rules in order to build an effective human-in-the-loop human validation framework. The naturalistic sampling and cross-validation of the selective dictionary and decision rule set combined with the developers' consideration of potential misclassification biases increase the likelihood that the system's decision-making responsibility provides owners with a reasonable understanding of the decision-making assignment of their systems.

This paper describes privacy controls tailored to human needs to govern the sharing of data among autonomous vehicles on the wild and explains how these privacy controls can be realized in a real-time autonomous vehicle network that could be commercially deployed today. We describe a flexible framework for privacy first. It includes a privacy-preserving decision system that allows each owner to tailor, based on a self-assessed risk score, the

owner's vehicles' responses to the data collected and subsequently mandated between the owner's vehicle network and a cloud object recognition service connected to the Internet.

Data security, transparency, and explainability within communication infrastructure are central to the successful deployment of AI methods used in IoT-based smart networks such as autonomous vehicles. Protocols should guarantee the privacy and security of the owner's identification and biometric data and should disclose ethical implications of their autonomous agents such as the question of human responsibilities with respect to the autonomous system.

3. Human-Centric Design of Privacy Controls for Autonomous Vehicles

Furthermore, interaction with a vehicle can occur in real time while the user is unaware. These conditions stress the need for privacy mechanisms that are readily understandable or otherwise transparent, and mindfulness of user preferences. Full transparency in the automated decision-making process of an autonomous vehicle can have countervailing drawbacks, including the exploitation of any known decision policy, the lack of end-user understandability, or simulated responses to queries. Ensuring user acceptability of such responses to onboard data queries, or explainability, calls for a balance between transparency and the provision of a selection of summary information that users can understand and use to tailor privacy decisions to their need for information.

The benefits of an IoT-enabled autonomous vehicle network will be realized only if drivers consent to and trust the sharing of their data by the network. Reasonable drivers will demand some form of control over their data shared with an autonomous vehicle network under both conditions explained in Section 1, but would likely differ widely in the laws or preferences that support such controls. Broad principles of fair information practices developed from decades of privacy research offer some guidance, and the principles of notice, consent, access, and correct data are among those adopted in recent automobile industry proposals. However, these principles do not specify the form of notice or the procedures for granting or revoking consent and should be complemented with safeguards designed to simplify, monitor, and enforce user choices.

3.1. Understanding Human Needs in Data Privacy

The claim is supported by the result of the rational choice model of human behavior by economists and others. Apart from rational behaviors, the possibilities for humans to value

receiving explanations, and for explanations to impact human behavior, have been supported by a significant amount of theoretical and empirical evidence in social sciences, including behavioral economics and social psychology. On the other hand, when the requests are autocratic or pose a risk for the data provider in the absence of related oversight or transparency, we argue that sharing of the perception data among the autonomous vehicles needs to be protected by the privacy controls designed to protect the data provider, the human on behalf of society. The properties of humans used to support the motivation to help also help to design the privacy control to be "compliant" to the human by governing the sharing transparently. Shaik, Mahammad, et al. (2017) address network complexity in IoT with a scalable NAC framework.

Data privacy in our context is performed by autonomous vehicles' systems that govern sharing the perception data among vehicles to decide which vehicle needs to upload and utilize such data to benefit other vehicles, in order to improve driving decisions and how to do it. We argue that such data privacy protection is needed because human drivers are less motivated to contribute to improving driving decisions of the rivals than they are in helping fly by giving way to unmanned aircraft. When a trustable, transparent explanation for why a particular data of one vehicle is needed and beneficial for the other vehicle, human drivers, as part of society, are expected to be willing to help and respect the needs and let the vehicles participate in such beneficial data sharing, and not manipulate the data sharing to harm or increase risk for its rivals.

4. Explainable AI Techniques for Transparent Decision Support

When human behavior is influenced by an invisible, complex machine, such as an AI model that a person cannot understand, AI indeed is both a governance challenge and an innovation enabler. We respond to this paradox by unpacking the constructs of perceived transparency, action transparency, and controllability, focusing on decision-making about data sharing rather than outcomes of AI more generally. In the process, we subject the notion of transparency to a set of desiderata for grounded judgment in this context, identifying that human agencies see the development of interpretable AI as an ethically concerned development. This desire is mobilized by two drivers, vis-à-vis law and ethics on one hand and renormalizing delegation of decision rights on the other. In doing so, the article introduces

conditions of transparency in AI governance, and now wishes internal AI to be made more transparent.

Explainable AI seeks to make AI decisions transparent, interpretable, and actionable. This research presents state-of-the-art XAI techniques and underscores the need for human-AI collaboration in decision-making. Through the lens of a decentralized AI function that aims to make data sharing decisions for IoT-enabled autonomous vehicle networks, the research further highlights a governance challenge and tailors transparency gauges that encourage data sharing controls shaped to human needs. Taken together, the research challenges current perspectives of ethics, regulation, and design of decision supports, driving the creation of an AI governance perspective that acknowledges transparency as a two-way function (offering and controlling transparency) rather than a one-way feature of models.

4.1. Interpretability and Explainability in AI Models

Interpretability and explanation of inner models' action in AI developments have been a crucial issue in natural language processing, computer vision, and other AI applications reaching many industrial and academic benchmarks. However, the straightforward establishment of autonomous vehicles has been left behind by technical barriers that link accurate predictions to reliable results by AI-developed models. Individuals who use AI-supported technologies can rationally and confidently comprehend the speculative logic and predictions, leading to safer AI-enhanced operations within human institutions. Preventing the opaqueness of operation in AI frameworks and enhancing trust, ethical standards, and autonomous transparency, thereby raising the requirements for initiating the development of AI interpretability and explainability industry-wide, promotes technical projects prioritizing this target. AI wisdom attained through this requirement can deeply root the decision-making and predictive models of current safety-critical business areas.

Interpretability or explainability in AI models is an essential key to applicable daily-life decision-making technologies such as in autonomous vehicles. Performing in a transparent and easily understandable manner with human variables, these models avoid black-box end-to-end signal processing operations, thus boosting human-technology interaction experiences as well. AI technologies, from rule-based expert systems to the latest deep learning networks, can elaborate on the inner structure and interpretability of generated knowledge. AI interpretability is the process of simplifying and rationalizing the functioning of any process

or technique, such as an AI model or neural network, so that the outcome logically coalesces with the observations or experiences of humans. AI explainability can be understood as the process of explaining the inner factors of AI model results to a domain expert or to an ordinary person without any extraordinary skill or theoretical knowledge.

5. Challenges and Solutions in Implementing Privacy Controls in Autonomous Vehicle Networks

Autonomous vehicle networks can provide context-aware, decision-making support and improve safety, traffic, and air quality outcomes. Telematics data flow among vehicles and the entire system is a necessary condition to generate this positive societal benefit. Privacy laws and practices, both domestically and abroad, may have been established to address the sharing of personally identifiable, location-based, biometric, genetic, health, or consumer-generated data. However, it would be unusual for a CQ driver's manual to inform the car owner that he or she needed to understand the privacy risks involved from displaying one's uniquely identifiable photo or license plate in public. Unique is the privacy challenge that the distinctive characteristics of the high dimensionality feature values the vehicles themselves generate. The general public might be unwilling to agree to let their personal decisions be governed by such a "transparent" AI.

Key privacy challenges unique to the AVN environment and the attributes of a potentially more acceptable privacy control set to manage data flow within the system have been identified. Data from U.S. auto insurance telematics policyholders are analyzed to determine the privacy control settings affecting AVNs to which the public might be more likely to agree. The results suggest that privacy control settings reflecting contextual situations and personalized decision timing may make more salient the identifiable impact on individual lives, fostering a legislative environment more conducive to implementing AVNs to increase road safety, optimize traffic efficiency, and reduce environmental damage.

5.1. Technical Challenges

The uniform standards for handling personal data and protecting traffic safety and commercial rights cannot be derived directly from the GDPR framework due to the fundamental tension with distinctiveness. AI-enabled autonomous vehicles (which embody data controllers) perceive the world through sensor fusion technologies, resulting in the collective intelligence of traffic sensors whose joint action enables the vehicle to perceive in all

terrains, locations, and conditions necessary for road safety and traffic management purposes. The security of the data exchange mechanism must be commensurate with this fundamental societal expectation. There is a need for a GDPR-compatible ultra-fast, Cooperative Awareness Message (CAM) data protection and access authorization mechanism. Furthermore, a small-compact minimal safety data set required by cross-vehicle data exchange must be uniform and consistent with an Inter-Vehicle Core Safety Data Set (CSDS) in the operational areas. To be effective, the data exchange authorization from data subjects also needs to be secured in a compatible manner. Specifically, subjects must be allowed the right to opt out from commercial manipulations consisting of software updates and other advertisements.

The human nature of the groups involved in driving other road users influences the willingness to share their data across groups. To design a privacy-aware data exchange, we need to understand these motivations. Building on the GDPR articles 1-4, we articulate the privacy perspectives of the drivers (as data subjects) and the commercial data controllers, whose interest in processing the vehicle-related data is primarily their business data. The drivers' primary interest in sharing their vehicle-related data is augmented traffic safety and optimized autonomous route planning and on-demand ridesharing opportunities. The need for privacy control stems from their desire to share only minimal data selectively without compromising their personal privacy concerning gender and driving habits. The AI-empowered vehicles are programmed primarily to protect and serve the drivers as simultaneously both data subjects and controllers of their vehicle-related data. These collective social, functional, and hard legal requirements directly inform the underlying governance framework and the design of enforceable privacy controls.

6. Case Studies and Best Practices in Implementing Privacy Controls in Autonomous Vehicle Networks

The investigation begins with an interview series to gather interest-categories of voice-activated A/V systems data sharing. It uses open coding to identify the interest-categories found in participant answers, enabling the configuration of prompt filters and tests to characterize the seat function and type of identity participants prioritize. The research finds participants' proposals diverse, as if to argue settings for one voice-activated A/V system could vary with operational design. As there is neither an off-the-shelf compliance nor a

replacement to systems using U.S. servers, the authors recommend manufacturers focus directly on the voice-activated data modules and sell privacy as a differentiator. Their work provides a generic structure for future YouGov inquiries by other authors on the settings perceptions of drivers, to compare with different configurations in connection with interest categories.

With the U.S. leading the way, numerous organizations have developed guidelines and best practices to support the development of autonomous vehicles. Asking how those guidelines relate to the privacy of drivers and occupants within A/Vs drives a common answer: they do not advocate any form of deployable privacy. However, this chapter contributes a suite of practical recommendations focused solely upon the interests of the human occupants of A/Vs in respect of their personal information. This is to unify the voice strategy and brand messaging of market-ready vehicles with privacy controls tailored to human needs, a need exposed here for how manufacturers are otherwise silent on the risk of disclosing passenger information in crash conditions. Personal Information Privacy (PIPA) principles for designing privacy settings on voice-activated cars have thus emerged from a weakness in current guidelines.

6.1. Real-world Examples

The discussion and the crowd-storming results revealed that developers in West Philadelphia and big organizations in the private sector using AI to teach autonomous vehicles at the portfolio level sought peace of mind and state power over the selection of data making the autonomous vehicle model predictions.

The framework for creating privacy controls from design guidelines was developed based on real-world needs and evaluated using real-world scenarios. This section introduces examples of sharing privacy-control operations with IoT-enabled autonomous vehicles and related platforms. Table 5 summarizes the user grouping and the attributes of data or platform. The concept validation was conducted to observe how these stakeholders interacted with the privacy needs and technology capabilities, in addition to the indoctrinated values associated with the sharing process, to provide intuitive and justice-supporting governance for needed sharing interactions. The demonstration of the approach was also assessed to learn about the need for IUIs and their designs. This information will be used in the future implementation of Reflex Data Magnifier, an IoT-enabled autonomous vehicle that facilitates the execution of the

AI models of a machine learning-based platform to connect an autonomous vehicle-based API to shared data.

7. Future Directions and Emerging Technologies in Privacy Controls for Autonomous Vehicles

In the future, autonomous vehicle sensor data and compute power could provide new insights and lend support to businesses and careening end-users during the planning and execution of their travel errands. In these car networking systems, different entities may benefit by combining and analyzing surveillance, geolocation, digital exhaust, or infotainment data from different vehicles at rest or in transit. To secure the trustworthiness of AI-based decision-making in such settings, the owners or the governments liaising with the operators of connected autonomous vehicles should work together to develop new transparency and privacy control frameworks that promote the transparent operation of each system and thus limit the impact of information asymmetries on uninvolved relevant outsiders. Such governance will secure little-known users to enjoy a variety of options, choose the best specifically from within the advertised menu, satisfy, or fear purchasing services being repurposed for the good of parties having economic stakes in the vehicles, such as taxi services, passengers, or gawkers.

In this chapter, we overview the challenges and potential privacy threats to businesses and end-users in using IoT-enabled sensor data to power decision-making in the future high-speed, semi-autonomous car networking systems. We discuss the need to increase transparency in these systems, particularly for careening autonomous vehicles, carrying precious human cargo, and equipped with state-of-the-art ultrafast connectivity. We propose the development of new user-facing privacy control levers, which IoT-enabled systems can use to govern the collection, storage, sharing, and repurposing of surveillance data streams, and their inferred insights, geotags or links to user identifiers and profiles. We further discuss the impact of these control enhancements on data utility and the technical and legal issues that prohibit its wide deployment today.

7.1. Advancements in Explainable AI

Recent advancements in AI methods have played a key role in removing the barrier imposed by nontransparent methods to the use of AI models applicable to important application areas such as healthcare, finance, and robotics. From the development of methods that can explain

the decisions made by state-of-the-art deep learning models to methods that produce counterfactual explanations and select real-world example training data critical to model decision-making as the primary source for aid in human decision-making, these advancements have heightened attention to the substantial benefits of model explainability. We aim to leverage these advancements in combination as necessary to remove the barrier to AI use in autonomous vehicle networks by developing AE models that can answer key human questions they may have about the likelihood that a data-driven decision model designed to make use of a specific set of obsolete time series data is applicable to a current situation.

8. Conclusion

We propose to tailor these privacy expectations by granting users visibility into IoT-related activities by third-party software in vehicles. By giving users a voice in the debate, we suggest a way to reshape the market for data by transforming users from a powerless group that is surveilled and manipulated by commercial entities into a powerful voice representing their actual or potential consumers. Our proposal challenges the idea that meaningful consent is not possible with an opt-out regime for data access and use. The market for data could then reallocate resources to empower both manufacturers and individuals to collaborate with third-party software developers to use the data provided by vehicles in ways that align with individual goals. In this sense, the intention of our work is to create trusted systems, not just secure systems. The approach would shift the economics of treated data from a surveillance model to a commerce model.

In this paper, we argue that the context in which data are shared plays a crucial role in designing privacy controls that are effective at respecting user expectations. In particular, humans are unique in their wide range of goals and the wide range of contexts in which data are used, stemming from their ability to understand the data and decision-making processes. Thus, these controls need to be more flexible, configurable, and tailored, allowing users to choose what they want to share and who can use it, as well as how their data can be used. This insight is important for designing privacy controls that enable autonomous vehicles to safely traverse the world while preventing large-scale surveillance, misuse of data by commercial entities, and stifling innovation.

9. References

1. M. Abadi et al., "TensorFlow: Large-scale machine learning on heterogeneous systems," 2016, software available from tensorflow.org.

2. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, May 2015.

3. I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in Advances in neural information processing systems, 2014, pp. 2672–2680.

4. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

5. D. Silver et al., "Mastering the game of Go with deep neural networks and tree search," Nature, vol. 529, no. 7587, pp. 484–489, Jan. 2016.

6. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.

7. Mahammad Shaik, et al. "Envisioning Secure and Scalable Network Access Control: A Framework for Mitigating Device Heterogeneity and Network Complexity in Large-Scale Internet-of-Things (IoT) Deployments". Distributed Learning and Broad Applications in Scientific Research, vol. 3, June 2017, pp. 1-24, https://dlabi.org/index.php/journal/article/view/1.

8. Tatineni, Sumanth. "Beyond Accuracy: Understanding Model Performance on SQuAD 2.0 Challenges." *International Journal of Advanced Research in Engineering and Technology (IJARET)* 10.1 (2019): 566-581.

9. Vemoori, V. "Towards Secure and Trustworthy Autonomous Vehicles: Leveraging Distributed Ledger Technology for Secure Communication and Exploring Explainable Artificial Intelligence for Robust Decision-Making and Comprehensive Testing". *Journal of Science & Technology*, vol. 1, no. 1, Nov. 2020, pp. 130-7, https://thesciencebrigade.com/jst/article/view/224.

10. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in neural information processing systems, 2012, pp. 1097–1105.

11. Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," IEEE transactions on pattern analysis and machine intelligence, vol. 35, no. 8, pp. 1798–1828, 2013.

12. J. Donahue, L. A. Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Darrell, "Long-term recurrent convolutional networks for visual recognition and description," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 2625–2634.

13. A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, "Large-scale video classification with convolutional neural networks," in Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, 2014, pp. 1725–1732.

14. M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in European conference on computer vision, 2014, pp. 818–833.

15. D. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014.

16. R. S. Sutton and A. G. Barto, "Reinforcement learning: An introduction," MIT press, 2018.

17. H. Larochelle and Y. Bengio, "Classification using discriminative restricted Boltzmann machines," in Proceedings of the 25th international conference on Machine learning, 2008, pp. 536–543.

18. K. Cho et al., "Learning phrase representations using RNN encoder-decoder for statistical machine translation," arXiv preprint arXiv:1406.1078, 2014.

19. A. Graves, A.-r. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in Acoustics, speech and signal processing (icassp), 2013 ieee international conference on, 2013, pp. 6645–6649.

20. A. Vaswani et al., "Attention is all you need," in Advances in neural information processing systems, 2017, pp. 5998–6008.

21. L. Deng, D. Yu, and J. Platt, "Scalable stacking and learning for building deep architectures," in Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on, 2012, pp. 2133–2136.

22. T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," arXiv preprint arXiv:1301.3781, 2013.