Neuro-Symbolic Reasoning in the Traffic Domain

Leilani H. Gilpin [a] and Filip Ilievski [b]

[a] *Department of Computer Science and Engineering, University of California, Santa Cruz, CA, United States*
E-mail: lgilpin@ucsc.edu

[b] *Information Sciences Institute, University of Southern California, CA, United States*
E-mail: ilievski@isi.edu

Abstract. Combining neural and symbolic features in a single *Neuro-Symbolic (NeSy)* AI reasoning system has shown the promise to bring the best of both worlds, yielding higher robustness and explainability. Yet, while NeSy reasoning has been shown to be effective on various tasks, its strengths and weaknesses for the traffic domain have not been systematically explored. To bridge this gap, we consider the promises and the challenges of NeSy reasoning in the traffic domain. We devise a taxonomy of traffic tasks that is organized into three categories: safety, perception, and diagnostic/complex inference. We consider these tasks from two perspectives: autonomous vehicles, where the goal is to minimize navigation errors of vehicle participants in traffic, and traffic monitoring, where the goal is to understand situations and their implications from a perspective of a third-person view. We investigate the role of NeSy reasoning for these tasks, aiming to understand its role in connecting different modalities, providing coherent explanations, supporting meaningful evaluation, and facilitating generalization to novel situations in the open domains. We consider the role of different knowledge types, including traffic rules, commonsense expectations, and causal links, in facilitating robust and explainable AI agents. We conclude with a list of open research directions geared towards reliable NeSy reasoning for traffic, tailored to social good and human-AI collaboration.

Keywords: neuro-symbolic reasoning, explainable AI, commonsense knowledge, intelligent traffic monitoring, autonomous driving, monitoring ML systems

1. Introduction

Developing reliable intelligent agents for the traffic domain has been an attractive pursuit due to the high-stake nature of this domain and the magnitude of its market [1]. While intelligent traffic agents leveraging the latest neural or symbolic advancements perform well in the lab, real-world usage has been marked with notorious examples that show that their reliability is far from the desired mark. A notable example of autonomous driving is the Uber fatality [2], which occurred due to the system focusing on a single modality (the perception system), rather than fusing multiple modalities of sensory information. Besides fusing multi-sensory information, the Uber accident shows a larger issue: autonomous vehicle may not know how to deal with inconsistent or unknown situations. For example, if an autonomous vehicle perceives an unknown object or faces a novel situation (e.g., an unknown traffic signal, like a flashing yellow light [3]), it may react in unpredictable ways. It may choose to ignore unknown objects [2] or extrapolate novel situations to prior experiences based on its underlying pattern-matching mechanisms. An example of the latter behavior is the accident of the Google Autonomous Vehicle (AV) in 2016, caused by the AV switching lanes after detecting sand bags.[1] Similarly, considering the related task of traffic monitoring, a processing

[1]https://www.wired.com/2016/02/googles-self-driving-car-may-caused-first-crash/, accessed on July 12, 2023.

system may detect that a vehicle and a bike lane overlap, but it would not be able to deduce that this represents a traffic violation. Moreover, the intelligent system may fail to understand certain context-specific behaviors [4] that happen in particular locations or at particular times. These experiences highlight the imminent need for reliable systems that are robust, explainable, and responsible. Merely relying on gathering sufficient data for each of these specific scenarios is unrealistic [5], thus emphasizing the need for novel approaches.

Meanwhile, Neuro-Symbolic (NeSy) reasoning, or the integration of neural and symbolic reasoning has become a staple of artificial intelligence and is often considered essential for robust intelligence in high-stake domains [6]. Given the high stake nature of the traffic domain, NeSy robust intelligence is a necessity for autonomous vehicles and traffic monitoring systems that are being entrusted with human-level decision-making. We see NeSy as a holistic platform to address these challenges in the traffic domain. By leveraging NeSy reasoning in both ego-centric autonomous driving and objective traffic monitoring, the traffic domain can benefit from advanced decision-making capabilities, enabling safer and more efficient transportation systems. NeSy addresses the knowledge integration problem of neural networks by supporting the fusion and abstraction of diverse information sources [7, 8]. This is a hallmark task for autonomous driving and the traffic domain where there are different types of information, including those from sensors, vision, and log data. In fact, there are autonomous driving data sets that have this type of multimodal data [9, 10]. By synthesizing such diverse and complementary types of information, symbolic reasoning enables them to be associated with domain-specific knowledge and rules, such as traffic regulations and traffic flow principles. It also enables integration with commonsense knowledge that may enable the models to generalize better to novel situations [11, 12]. NeSy reasoning provides an opening for the resulting models to be *explainable*, by transforming the explicit knowledge into human-readable format; *adaptive*, by leveraging the knowledge to transfer better to novel situations; *responsible*, by providing a mechanism for communication with the model in a meaningful way; and *collaborative*, by facilitating the extraction and representation of meaningful information [13]. Yet, while NeSy reasoning has been shown to be effective in various domains, its strengths and weaknesses for the traffic domain have not been systematically explored.

In this paper, we investigate the role of NeSy reasoning in addressing the complex challenges present in the traffic domain. We first provide background about prior work on monitoring autonomous vehicles and relevant neuro-symbolic methods (section 2). We focus on two core subproblems that stem from complementary perspectives: autonomous driving (first-person view) and traffic monitoring (third-person view) (section 3). We show common failures in both subproblems, and discuss the challenges and possible solutions highlighting the need for NeSy representations and reasoning. We consider the following symbolic knowledge within broader NeSy solutions: commonsense knowledge, rules, soft constraints, and causal reasoning. We describe their role to support better robustness on novel tasks and explainability to relevant users (section 4). We conclude with a list of challenges that motivate future work for NeSy reasoning in the traffic domain (section 5).

2. Background

Our paper examines NeSy challenges in the traffic domain. In the following sections, we describe the current state of the subtasks: autonomous vehicles and traffic monitoring, and we provide background on knowledge graphs. Namely, for the remainder of the paper, we will examine traffic domain tasks from two perspectives. Autonomous vehicles will be in the "first person view", meaning we are considering one autonomous vehicle at a time, typically from the perspective of the AV itself. When we talk about a larger world view, e.g., the "third person view," that means that we are considering traffic situations from an objective standpoint, e.g., a stationary camera.

We provide background on common representation and aggregation methods in the form of knowledge graphs to explore the synthesis of NeSy approaches and their potential application in traffic domain challenges. We do not focus on novel neural architectures for end-to-end training of autonomous vehicles and traffic monitoring systems [14], as these are consistently changing and have been covered well in recent survey papers for first-person driving [15], object recognition [16], and for safe driving [17], as well as broad surveys on NeSy methods and challenges [18–20].

2.1. The Current State of Autonomous Vehicles

Although there have been optimistic promises of fully autonomous vehicles in 5-10 years,[2] autonomous vehicles learn and drive differently than humans do. Autonomous vehicles are susceptible to failures, especially out-of-domain errors. There are many ways to "fool" an autonomous vehicle [21], and the failure cases cannot be enumerated.[3] There have been developments towards test suites of plausible failure modes [24], but creating more accurate failure detection techniques is not enough. The underlying systems need the capability to *introspect* about their own behavior and learn through experience. For example, in the Uber self-driving accident, a software system within the AV ignored the pedestrian detected by the LiDAR sensor to be a false positive detection, resulting in a pedestrian fatality [2]. The software system lacked the common sense to know that an object moving in the middle of the road is likely a pedestrian. The autonomous vehicle should *learn* from this mistake and ensure it will not be a repeated failure case. NeSy reasoning can be used to integrate symbolic rules with multimodal vehicle and traffic data.

Human drivers are also susceptible to failures. But human drivers fail differently than autonomous drivers. Human drivers are able to adapt to driving in new environments. Humans learn new driving rules by examining the context and reflecting on their behavior: they explain the best action or its alternative to themselves. Take, for example, the behavior of "flashing high beams." Based on the location, context, and social rules, we may apply an existing rule (if a neighboring car is flashing its high beams, then turn on the vehicle's headlights), or we may create a new rule based on a new explanation (e.g., since they are flashing their high beams aggressively, that means to watch out for something). NeSy enables systems to continuously adapt and refine their knowledge based on real-world interactions.

2.2. The Current State of Traffic Monitoring

Intelligent traffic monitoring has a wide range of use cases, which typically assume the existence of video footage from a stationary surveillance camera. The intelligent system needs to be able to understand the elements of a scene in a given moment, understand whether it depicts an anomalous configuration (e.g., an accident or some other event), and perform diagnostics and complex inference through tasks like introspection and forecasting. Most of the state-of-the-art methods have focused on perceptual tasks (e.g., identifying the number of lanes and segmenting the traffic participants in a scene), whereas more complex reasoning tasks have received little attention. As a result, there has been progress in terms of perceptual and simpler classification tasks, whereas more extensive reasoning remains out of the reach of current technology. The first causal reasoning task for traffic monitoring has been BDD [25], followed by the TrafficQA [26] task that requires six forms of complex inference, and by the suite of knowledge-intensive textual inference tasks proposed by Zhang et al. [5]. The reason behind this discrepancy is that techniques for causal inference and commonsense reasoning have not been applied to traffic monitoring, and learning complex inference from data directly is non-trivial [5, 26].

2.3. Knowledge Graphs for Aggregating Symbolic Knowledge

Knowledge graphs can provide a comprehensive representation of the complex factors that influence traffic safety. This allows us to identify patterns, correlations, and potential risks that would otherwise be difficult to discern. Reasonable knowledge can be provided to the monitoring system from a commonsense knowledge base. In the adaptable monitoring system, this knowledge is parsed into a web standard. Commonsense knowledge bases are key tools for developing systems that understand natural language descriptions and produce explanations. CYC is regarded as the world's longest-lived artificial intelligence project [27], with a comprehensive ontology and knowledge base with basic concepts and "commonsense rules," but there have been significant challenges to using CYC for real-world applications in natural language processing (NLP) and computer vision (CV) [28], including its proprietary nature and the difficulty of grounding of situations to CYC. Speer and Havasi [29] contributed ConceptNet5, a

[2] In May 2018, Andrew Ng said that autonomous vehicles are "here": https://medium.com/@andrewng/self-driving-cars-are-here-aea1752b1ad0

[3] Some examples are security hacks [22], or adversarial attacks [23]

Table 1

Example tasks for first (1st) person view (autonomous vehicles) and third (3rd) person view (traffic monitoring) that demonstrate the need NeSy Representations and Reasoning.

Task Classification	View	Sample Challenge
Safety	1st	Pass a dynamic driving test for autonomous vehicles.
Safety	3rd	Understanding of rare road events, e.g., extreme weather.
Perceptual	1st	Classify out-of-domain objects in the AV's vicinity.
Perceptual	3rd	Understanding elements of traffic scenes based on perception (e.g., number of vehicles).
Inference	1st	Introspection: sensory failure, NeSy for reliance on secondary modalities
Inference	3rd	Establishing links between observed events and their likely causes.

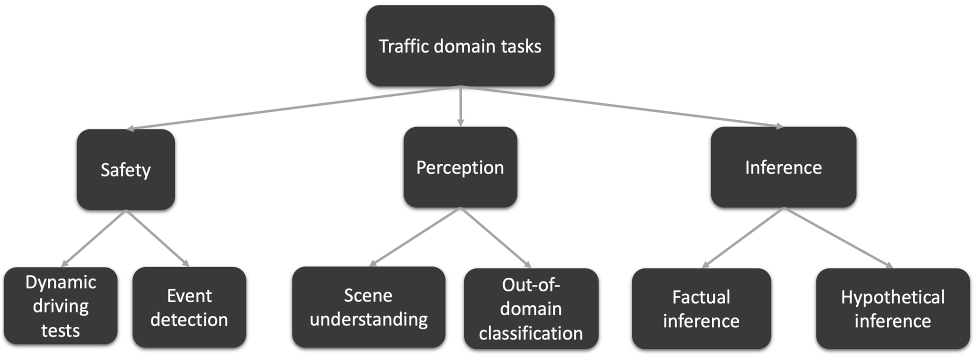

Fig. 1. Our taxonomy of traffic tasks: safety, perceptual, and inference.

freely-available semantic network of popular commonsense knowledge. The combination of popular commonsense knowledge sources into a single graph, as done within the CommonSense Knowledge Graph [30] and NextKB [31] provides an opportunity for richer and more comprehensive knowledge to be used for reasoning in traffic. The aggregated commonsense knowledge can be further organized into high-level knowledge dimensions [32] and aligned with commonsense axioms [33]. Meanwhile, domain knowledge can be distilled from driving manuals to improve the model's understanding of traffic rules and situational constraints. Combining these complementary kinds of knowledge is expected to enhance the comprehension of ambiguous or contextually rich language by providing contextual cues and background information [5].

3. Traffic Understanding Needs NeSy Representations and Reasoning

Traffic understanding is a complex multimodal challenge consisting of object detection and classification, object localization, trajectory prediction, sensor fusion, and planning. Traffic understanding encapsulates a unique blend of low-level perception tasks [15] and human-level symbolic reasoning and cognition [34], making it a neuro-symbolic goal. Curiously, despite the importance of reliable traffic understanding, there is no clear organization of the tasks within this domain. We bridge this gap by organizing the traffic understanding tasks into a taxonomy in Figure 1. Our taxonomy has three categorizations of tasks: **Safety** tasks (in Section 3.1), **Perceptual** tasks (in Section 3.2), and **Inference** tasks (in Section 3.3). Each of these task categories has been considered from an ego-centric perspective of an autonomous vehicle (first-person view) and from a perspective of a traffic observer like a camera (third-person view), resulting in corresponding variants. Table 1 shows first- and third-person view examples for each of the three traffic task categories. We next review the tasks within each category in turn.

3.1. Safety

Safety in traffic domain tasks is of the utmost importance, as traffic fatalities are increasing,[4] and as of recently, traffic fatalities are at a 16-year high.[5] One of the promises of intelligent traffic monitoring and autonomous vehicles is to reduce driving fatalities and develop driver assistance technologies, which could mitigate injury and harm. However, safety in the traffic domain requires real-time analysis (data-driven neural processing) and decision-making (symbolic reasoning), making it suitable for the integration of neural and symbolic techniques. For example, consider a traffic monitoring system that is reporting on current traffic conditions, but it is unreliable in weather conditions like rain, sleet, and snow. Neural networks may struggle to learn patterns in these situations, and incorporating symbolic reasoning can help handle mitigate and explain these extreme situations.

One challenge for autonomous vehicles is to create **dynamic driving tests**, which mimic how human drivers are evaluated as safe or unsafe drivers. For example, when humans learn to drive, we start by memorizing a set of safe driving rules from a driving handbook. Recent work has learned these safe driving rules from manuals [35], in order to codify and standardize these behaviors in a common symbolic language. Generative models, such as Generative Adversarial Networks (GANs) [36] or Variational Autoencoders (VAEs) [37], have the ability to learn the underlying distribution of driving data and generate new samples that resemble real-world driving situations. These generative models can be combined with a symbolic system to create dynamic "dangerous" driving conditions [38], e.g., construction zones, sporadic pedestrian crossings, etc. Previous work on reasonableness monitors [39, 40], checked and validated reasonableness with a production-level reasoner [41]. It is essential to use these generated tests in conjunction with real-world testing and validation, including diagnostics (see Section 3.3), to ensure the safety and reliability of autonomous driving systems. Driving tests have recently been introduced from an objective monitoring perspective, prompting agents to select the legal or reasonable course of action in a given situation [5]. Namely, the HDT-QA benchmark tests intelligent agents on human-driving tests for three driver categories (Motorcycle, Car, and Commercial Driver) in each of the 50 US states.

Another task for ensuring safety in both AVs and traffic monitoring scenarios is **event detection**, i.e., understanding anomalous situations and extreme environments. Datasets and challenges for event detection should include various road types, traffic conditions, weather conditions, pedestrian interactions, and other potential hazards. Intelligent traffic monitoring datasets have initially focused on detecting the occurrence of a certain event (e.g., an accident). DeepRacer is an educational platform for autonomous vehicle racing, designed primarily to experiment with reinforcement learning methods in intelligent control systems [42]. ROAD: The Road Event Awareness Dataset for Autonomous Driving [43], has been designed to test the autonomous vehicle's understanding of road events. Its successor, ROAD-R, is a neuro-symbolic autonomous driving dataset that investigates whether the models remain safe and consistently follow the environmental constraints throughout their performance [44]. One possible approach for tasks like ROAD-R is to design and implement memory-efficient t-norm losses for the task of event detection in autonomous driving. CADP [45] is a spatiotemporally annotated dataset for accident forecasting using traffic camera views. The Berkeley Deep-Drive dataset (BDD) contains real driving videos containing abundant driving scenarios [25], where the task of the agent is to recognize the action in the video (e.g., the vehicle is accelerating). There have been many other datasets for anomaly detection in road traffic using visual surveillance, surveyed in [46]. Event detection in traffic is a crucial aspect that helps identify potential hazards and inform decision-making processes, but addressing the challenges faced by AVs and traffic monitors requires a comprehensive approach encompassing various safety, technological, and regulatory considerations.

3.2. Perception

The fusion of neural networks and symbolic reasoning plays a crucial role in machine perception. Perception tasks have been a challenge of NeSy reasoning over AI history, from early machine vision systems [47, 48] to cutting-edge standardized models of human cognition [49]. In the traffic domain, perception tasks are the primary task:

[4] In CA, traffic fatalities increased from "approximately 7.6% from 3,980 in 2020 to 4,285 in 2021." via https://www.ots.ca.gov/ots-and-traffic-safety/score-card/

[5] Via https://www.nhtsa.gov/press-releases/early-estimate-2021-traffic-fatalities

recognizing objects, pedestrians, road signs, traffic lights, and other vehicles. While many traffic systems use end-to-end learning [14], we often require additional (symbolic) reasoning to make trustworthy decisions. For example, consider driving in a new rural environment. The vehicle may encounter an object it has never seen before: a fallen tree branch. Since the vehicle has not seen this object before, it chooses to ignore it, causing the vehicle to run over the branch, which gets stuck in its undercarriage, preventing the vehicle from moving properly.

Scene understanding can be seen as an umbrella task that encapsulates many of the visual perception tasks into a common framework. Several datasets focus on a fine-grained understanding of the environment, such as object segmentation [50, 51] and scene entity prediction [52]. The Stanford Cars dataset [53] contains 3D object representations for fine-grained categorization. The VERI-Wild [54] dataset evaluates the ability of vision models to re-identify vehicles in surveillance camera footage. The related Stanford Drone Dataset [55] tests systems' ability to predict human trajectories in crowded scenes. TrafficQA [26] consists of over 60K QA samples based on over 10k traffic scenes. Most of the questions in this dataset belong to the category of basic understanding, e.g., how many lanes are there on the freeway? The successful completion of these traffic understanding tasks lays the foundation for addressing the out-of-domain challenges of autonomous vehicle perception.

The main challenge in traffic perception is handling **out-of-domain classification**. There are perception challenges for autonomous vehicles [9], but these are classic supervised machine learning tasks: with a training and test dataset. These do not characterize the unique challenges in perceiving new objects while driving. For example, consider an autonomous vehicle driving through a new neighborhood. The perception sees a skateboard and correctly identifies the object. The problem is that the vehicle has no symbolic knowledge or context. As humans, we know that if we see a skateboard in a neighborhood, then many children are likely following the skateboard. Other challenges include adversarial attacks, where a few pieces of tape can fool the perception system into thinking a stop sign is a 45-mph sign [56]. One way to mitigate perception challenges in intelligent traffic monitoring is to develop representative datasets and understand the relation between task properties and the properties of the task [5]. Possible NeSy solutions include adding contextual symbolic information, e.g., the shape, size, and color of the sign, or designing rule-based stress tests to find the point of failure [57]. Incorporating NeSy approaches into perception systems for autonomous vehicles can create safer, more reliable vehicles in real-world environments.

3.3. *Inference*

Many realistic tasks require higher-order inference of factual or hypothetical aspects of a situation. Such inference is analogous to model-based diagnostics in computer science [58, 59], which refers to the process of identifying and analyzing problems, errors, software faults [60], or issues in complex computer systems. The goal of diagnostics for AVs is to understand the root cause of a problem or malfunction and provide insights into how to resolve it effectively, similar to the role of a highway patrol officer to patrol for unsafe driving conditions, or a mechanic to monitor and maintain vehicles. A related task from a traffic monitoring perspective is simulating situations forward or backward to enhance the model's understanding of a situation. The quest for incorporating inference skills in traffic agents represents a shift in paradigm, from the dominant focus on perception to *understanding*, which requires abundant domain knowledge and commonsense reasoning methods.

In the AV domain, the task of introspective diagnostics has been popular, covering both **factual and hypothetical reasoning**. For example, consider an introspective case of self-diagnosis where an autonomous vehicle is consistently driving well on the freeway. When alerts or diagnoses appear, the vehicle can self-diagnose and repair the problems without issues. However, the autonomous vehicle self-diagnosis system may have limitations in nuanced situations like construction zones or missing lane markers, leading to potential errors that go unfixed or ignored. For example, if a self-driving car's sensory system fails due to weather [61], how can the vehicle rely on other sensory modalities to safely continue its journey? In this case, it is crucial to have a comprehensive understanding of an autonomous system, which a model-based system can represent. Autonomous vehicles claim to be "self-diagnosing" [62, 63], they are able to find the root cause of their own errors and fix it. However, there is still a need for model-based, or symbolic reasoning for, e.g., reasoning about plans [64] and hypothetical reasoning [65]. By incorporating data-driven diagnosis with a model-based (symbolic) diagnostic system, users can be more confident in the AV system's trustworthiness and reliability.

Table 2
Examples of BDD-QA, TV-QA, and HDT-QA from [5]. (*) denotes the correct answer.

BDD-QA
Q: The car in front of the car is slow, but the traffic is also heavy in other lanes, what will the car do next?
A1: The car speeds up and turns to the right; **A2:** The car moves back to the right side of the road; **A3:** The car slows down(*);
A4: The car backs up slowly
TV-QA
Description: The POV car is quickly going down a highway. The POV car approaches an intersection. There is a red sedan in the opposing lane waiting to turn and cross the intersection. The red sedan quickly makes a left turn. when the POV car enters the intersection. The POV car veers to the right. The red sedan hits the side of the POV car.
Q: Could the accident be prevented if the involved vehicles change lane or turn properly?
A1: Yes(*); **A2:** No, that was not the main cause of the accident
HDT-QA
Q: If you find yourself in a skid:
A1: Brake lightly; **A2:** Brake abruptly; **A3:** Stay off the brakes(*)

From a traffic monitoring perspective, tasks relating to both **factual and hypothetical reasoning** have been considered either separately or jointly [5, 26, 66]. The BDD-X dataset [66] enhances the popular BDD dataset consisting of action descriptions, by adding explanations (e.g., the vehicle accelerated because the traffic light was green). This dataset was reused once more to create BDD-QA [5], a causal reasoning QA dataset generated based on careful engineering over the descriptions of actions and their explanations. BDD-QA can be seen as a decision-making dataset for predicting the best course of action, where the actions can be grouped into seven categories: accelerate, merge, drive, slow, stop, turn, and navigate. The QA set of TrafficQA [26] includes six different aspects of reasoning problems, five of which can be categorized as inference tasks: attribution, introspection, counterfactual inference, event forecasting, and reverse reasoning. TV-QA [5] extended TrafficQA with detailed captions for complex actual and hypothetical reasoning over traffic situations in the text on four tasks: reverse reasoning, forecasting, counterfactual reasoning, and introspection. An example of the three datasets from this work is shown in Table 2. By leveraging these datasets, we can make advancements in traffic monitoring for enhancing system reliability.

4. NeSy Methods for Traffic Understanding

We expect that NeSy reasoning methods can enable reliable technology for traffic understanding. Next, we review prior methods that contribute to this goal along two complementary aspects: **robustness** and **explainability**.

4.1. NeSy for Robust Traffic Understanding

NeSy can leverage background symbolic knowledge in a neural architecture to glue together information from various domains, modalities, and sources. By leveraging the wealth of information encapsulated in these knowledge bases, it becomes possible to bridge the gap between diverse modalities such as text, images, audio, and video. By using knowledge to enhance traffic understanding, the authors in [67] create and evaluate knowledge graph embeddings for autonomous driving. [52] enhance the knowledge graph for scene entity prediction in autonomous driving. ITSKG [68] is a knowledge graph framework for extracting actionable information from raw sensor data in traffic. CoSI [69] is a knowledge graph-based approach for representing information sources relevant to traffic situations. Such works focus on in-domain reasoning and rely on benchmark-specific training data.

Meanwhile, the robustness of methods can be tested through a zero-shot evaluation procedure. In [5], we used a natural language formulation of traffic reasoning and experimented with equipping a language model with domain knowledge from a QA benchmark [70], commonsense knowledge from a synthetically created QA set [71], and their combination. The resulting model was evaluated on an unseen test set that evaluates causal inference, on which the aggregated knowledge performed the best, the vanilla model was the worst, and commonsense knowledge was more beneficial than domain knowledge. The contribution of various knowledge types was task-dependent: commonsense knowledge is most impactful for complex and hypothetical tasks, like introspection, whereas retrieving domain knowledge is most effective for tasks that require contextual decision-making and understanding of traffic rules [5].

To illustrate the complementarity of these two knowledge types, let us consider the question *what might be the reason that a car is waiting in the intersection when the traffic light is green?* Its answer *The car is waiting for pedestrians* is well-supported by both kinds of knowledge: commonsense knowledge can tell language models that cars will pass the crosswalk when driving and crosswalk will appear at the intersection, while traffic domain knowledge tells the models that cars should yield to pedestrians passing the crosswalk [5].

These findings suggest that by incorporating richer knowledge into the modeling process, we can unlock new opportunities for the development of intelligent systems capable of reasoning across different modalities and delivering more comprehensive and accurate results. Yet, we are only beginning to realize this potential. To this end, HANS [1] provides a theoretical neuro-symbolic framework for integrating different modalities and knowledge types into a single neural system, consisting of six general processes: generation, semantic representation, augmentation, assessment, infusion, and inference. Yet, the combination of multiple modalities and knowledge types for the traffic domain remains an open challenge. The idea of scene knowledge graphs (SKGs) [72] can intuitively be used as a common representation, supported by traffic monitoring formalisms like the Scene Ontology [73] and the Traffic Monitoring Knowledge Graph [1]. Following [72], distant supervision data can be extracted to transform situations into such SKGs, and the SKGs into decisions or classification outputs. These are open-ended research directions that require significant innovation both for the traffic domain and multimodal reasoning in general.

4.2. NeSy for Explainable Traffic Understanding

Commonsense knowledge bases provide a structured and organized representation of human knowledge. This knowledge can be used to explain how and why certain decisions or actions are taken. For example, if a decision is made by a machine learning model based on commonsense knowledge, it can be explained by referring to the relevant facts, rules, and principles stored in the knowledge base. This makes it easier to understand the decision-making process and to identify any biases or errors that may have occurred.

The integration of symbolic reasoning allows for the explicit representation of knowledge in the form of rules, leading to an understandable explanation. In the context of autonomous driving, these rules are essential to abide by the "rules of the road." By incorporating symbolic representations into the reasoning process, neuro-symbolic systems can provide explanations that are grounded in human-understandable concepts. For example, if an autonomous vehicle is unsure of what it is perceiving, instead of showing a saliency map [74] on possibly irrelevant parts of the input image, we can construct an interpretable, natural language explanation showing that the vehicle was unsure of what it perceived. These symbolic representations can be used to represent domain-specific knowledge, causal relationships, and decision-making rules, making the reasoning process more transparent and interpretable.

By combining neural networks and symbolic reasoning, NeSy reasoning can bridge the gap between opaque deep learning models and symbolic-level human understanding. This enables explanations that are more transparent, interpretable, and dynamic. For example, consider our prior work on a multimodal monitoring system for autonomous vehicles [75]. This system adds a set of domain specific rules and commonsense knowledge to explain the failures between parts, e.g., when the vision system and sensor system disagree on what is being perceived. These explanations also reflect (and explain) the inherent multimodal nature of autonomous systems and traffic. Neural networks learn from complex, high-dimensional data, and symbolic reasoning, like diagnostics brings transparency and interpretability. By integrating these two approaches, neuro-symbolic systems can leverage the strengths of each. This also aligns with the need for explanations for accountability, so that we can diagnose and understand the errors in complex systems.

5. Conclusions and Outlook

In this position paper, we examined NeSy representation and reasoning in the context of the traffic domain. We decomposed the traffic domain into two views: first-person (autonomous vehicles) and third-person (traffic monitoring) and contributed a taxonomy for traffic domain tasks: safety, perception, and inference tasks. We provided examples of the tasks in both views and discussed the prior work on them. Throughout the paper, we show that NeSy plays a pivotal role in creating robust and trustworthy traffic systems, creating a more reliable transportation infrastructure for society as a whole. Specifically, we discussed prior work that combines neural and symbolic techniques to improve two aspects of the methods: robustness and explainability. The analysis in this paper highlights that the traffic domain covers a rich set of tasks, and relies on difficult, open-world research problems that we are only starting to understand and address as a community.

Research on developing AI for the traffic domain has gradually been switching its attention to NeSy evaluation, representation, and reasoning methods, in recognition of the requirements of affected users. We believe that reliable intelligent agents for traffic need to satisfy the CARE (Controllable, Adaptive, Responsible, and Explainable) [13] principles in order to make a positive impact on people and society as a whole. Yet, our paper shows that there is significant space for improvement of the current AI technology for the traffic domain. While addressing the task challenges from our taxonomy for the traffic domain is crucial, these efforts also open research directions in NeSy in the traffic domain. Here, we list three key research directions, posed as requirements, to motivate reliable NeSy reasoning for real-world traffic situations:

1. *Representative evaluations for traffic systems* need to be multimodal, interactive, and realistic. While there are several autonomous driving challenges, including multimodal task challenges,[6] there is a lack of neuro-symbolic challenges with these properties. The more recent challenges in trajectory prediction are hybrid tasks,[7] which involve navigating through dynamic and unpredictable environments with a wide range of potential interactions and uncertainties. Upgrading these tasks to more realistic settings that cover the entire complexity of the safety, perception, and inference tasks is a notable future work goal. We expect that NeSy will be a great asset on such tasks because it gives us the tools, e.g., rules, KBs, and reasoning to address the complexities and uncertainties in real-world traffic applications.
2. *Comprehensive knowledge for traffic AI models* must contain a high-quality combination of ontological concepts, rules, domain knowledge, and commonsense knowledge. Recent work has shown that such knowledge types provide complementary insights [5]. We expect that comprehensive sources of knowledge or methodologies for combining knowledge on the fly in light of the context will be essential for AI that pertains to the CARE principles and performs reliably. We expect that NeSy methods are uniquely positioned to contribute to the development of such resources and their usage for downstream reasoning.
3. *NeSy user-facing models* must be robust, explainable, and controllable by humans in intuitive ways. Both autonomous driving and traffic monitoring are sensitive tasks that require intelligent agents that not only perform well, but also perform responsibly, can explain their predictions, and can interact with people over time. This goal is at the forefront of NeSy AI for a variety of applications; however, the traffic domain is especially attractive given its high stakes and profitability.

Embracing these principles can lead to advancements in applying intelligent agents in the traffic domain, enabling safer, more efficient, and reliable traffic systems. For example, consider the third person's view of stop lights or traffic control systems. Traditional traffic lights follow a fixed schedule, leading to inefficiencies and congestion, especially during peak hours or in response to unexpected events. Instead, let us consider an AI-powered traffic light, with NeSy representation and reasoning. The control system can dynamically adapt signal timings based on real-time traffic conditions, reducing congestion and improving the overall flow of vehicles. It is also explainable, and it can signal the reason why the light is taking so long. This AI-enabled traffic light would also have an impact beyond saving time and frustration. In cases of an emergency, the traffic light could quickly adjust lights to clear a pathway for emergency vehicles, helping to lives. In conclusion, NeSy reasoning is imperative to safety in the traffic domain; thus, we encourage the development of NeSy challenges and methods to reduce traffic accidents, reduce congestion, save lives, and foster a culture of responsible driving.

[6]NuScenes challenges: https://www.nuscenes.org/object-detection?externalData=all&mapData=all&modalities=Any
[7]NuScenes Trajectory Prediction Challenge: https://www.nuscenes.org/prediction?externalData=all&mapData=all&modalities=Any

References

[1] E. Qasemi and A. Oltramari, Intelligent Traffic Monitoring with Hybrid AI, *arXiv preprint arXiv:2209.00448* (2022).
[2] A. Marshall and A. Davies, Uber's Self-Driving Car Saw the Woman It Killed, Report Says.
[3] T. Chen, Augmenting anomaly detection for autonomous vehicles with symbolic rules, Master's thesis, MIT, 2019.
[4] D. Ortiz, L.H. Gilpin and A.A. Cardena, Semi-Automated Synthesis of Driving Rules, *Symposium on Vehicle Security and Privacy (VehicleSec)* (2023).
[5] J. Zhang, F. Ilievski, A. Kollaa, J. Francis, K. Ma and A. Oltramari, A Study of Situational Reasoning for Traffic Understanding, in: *KDD*, 2023.
[6] G. Marcus, The next decade in AI: four steps towards robust artificial intelligence, *arXiv preprint arXiv:2002.06177* (2020).
[7] R. Manhaeve, S. Dumancic, A. Kimmig, T. Demeester and L. De Raedt, DeepProbLog: Neural Probabilistic Logic Programming, in: *Advances in Neural Information Processing Systems*, Vol. 31, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi and R. Garnett, eds, Curran Associates, Inc., 2018. https://proceedings.neurips.cc/paper_files/paper/2018/file/dc5d637ed5e62c36ecb73b654b05ba2a-Paper.pdf.
[8] J. Mao, C. Gan, P. Kohli, J.B. Tenenbaum and J. Wu, The Neuro-Symbolic Concept Learner: Interpreting Scenes, Words, and Sentences From Natural Supervision, in: *International Conference on Learning Representations*, 2019. https://openreview.net/forum?id=rJgMlhRctm.
[9] H. Caesar, V. Bankiti, A.H. Lang, S. Vora, V.E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan and O. Beijbom, nuScenes: A multimodal dataset for autonomous driving, *arXiv preprint arXiv:1903.11027* (2019).
[10] R. Kesten, M. Usman, J. Houston, T. Pandya, K. Nadhamuni, A. Ferreira, M. Yuan, B. Low, A. Jain, P. Ondruska, S. Omari, S. Shah, A. Kulkarni, A. Kazakova, C. Tao, L. Platinsky, W. Jiang and V. Shet, Lyft Level 5 AV Dataset 2019, 2019.
[11] K. Ma, F. Ilievski, J. Francis, Y. Bisk, E. Nyberg and A. Oltramari, Knowledge-driven Data Construction for Zero-shot Evaluation in Commonsense Question Answering, in: *AAAI*, 2021.
[12] P. Wang, F. Ilievski, M. Chen and X. Ren, Do language models perform generalizable commonsense inference?, *arXiv preprint arXiv:2106.11533* (2021).
[13] Z. Akata, D. Balliet, M. De Rijke, F. Dignum, V. Dignum, G. Eiben, A. Fokkens, D. Grossi, K. Hindriks, H. Hoos et al., A research agenda for hybrid intelligence: augmenting human intellect with collaborative, adaptive, responsible, and explainable artificial intelligence, *Computer* **53**(08) (2020), 18–28.
[14] M. Bojarski, D. Del Testa, D. Dworakowski, B. Firner, B. Flepp, P. Goyal, L.D. Jackel, M. Monfort, U. Muller, J. Zhang et al., End to end learning for self-driving cars, *arXiv preprint arXiv:1604.07316* (2016).
[15] S. Grigorescu, B. Trasnea, T. Cocias and G. Macesanu, A survey of deep learning techniques for autonomous driving, *Journal of Field Robotics* **37**(3) (2020), 362–386.
[16] H. Fujiyoshi, T. Hirakawa and T. Yamashita, Deep learning-based image recognition for autonomous driving, *IATSS research* **43**(4) (2019), 244–252.
[17] K. Muhammad, A. Ullah, J. Lloret, J. Del Ser and V.H.C. de Albuquerque, Deep learning for safe autonomous driving: Current challenges and future directions, *IEEE Transactions on Intelligent Transportation Systems* **22**(7) (2020), 4316–4336.
[18] U. Kursuncu, M. Gaur and A. Sheth, Knowledge infused learning (k-il): Towards deep incorporation of knowledge in deep learning, *arXiv preprint arXiv:1912.00512* (2019).
[19] A.d. Garcez and L.C. Lamb, Neurosymbolic AI: The 3rd wave, *Artificial Intelligence Review* (2023), 1–20.
[20] W. Wang and Y. Yang, Towards Data-and Knowledge-Driven Artificial Intelligence: A Survey on Neuro-Symbolic Computing, *arXiv preprint arXiv:2210.15889* (2022).
[21] S.E. Anthony, The Trollable Self-Driving Car, 2016, (Accessed on 07/01/2020).
[22] K. Koscher, A. Czeskis, F. Roesner, S. Patel, T. Kohno, S. Checkoway, D. McCoy, B. Kantor, D. Anderson, H. Shacham et al., Experimental security analysis of a modern automobile, in: *2010 IEEE symposium on security and privacy*, IEEE, 2010, pp. 447–462.
[23] K. Eykholt, I. Evtimov, E. Fernandes, B. Li, A. Rahmati, C. Xiao, A. Prakash, T. Kohno and D. Song, Robust physical-world attacks on deep learning visual classification, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 1625–1634.
[24] Y. Tian, K. Pei, S. Jana and B. Ray, DeepTest: Automated Testing of Deep-neural-network-driven Autonomous Cars, in: *Proceedings of the 40th International Conference on Software Engineering*, ICSE '18, ACM, New York, NY, USA, 2018, pp. 303–314. ISBN 978-1-4503-5638-1. doi:10.1145/3180155.3180220.
[25] H. Xu, Y. Gao, F. Yu and T. Darrell, End-to-end learning of driving models from large-scale video datasets, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2174–2182.
[26] L. Xu, H. Huang and J. Liu, Sutd-trafficqa: A question answering benchmark and an efficient network for video reasoning over traffic events, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 9878–9888.
[27] D.B. Lenat, R.V. Guha, K. Pittman, D. Pratt and M. Shepherd, CYC: Toward programs with common sense, *Communications of the ACM* **33**(8) (1990), 30–49.
[28] K. Mahesh, S. Nirenburg, J. Cowie and D. Farwell, An Assessment of CYC for Natural Language Processing, Memoranda in Computer and Cognitive Sciences MCCS-96-296, Computing Research Laboratory, New Mexico State University, Las Cruces, NM, 1996.
[29] R. Speer and C. Havasi, ConceptNet 5: A large semantic network for relational knowledge, in: *The People's Web Meets NLP*, Springer, New York, 2013, pp. 161–176.
[30] F. Ilievski, P. Szekely and B. Zhang, CSKG: The CommonSense Knowledge Graph, in: *Extended Semantic Web Conference (ESWC)*, 2021.

[31] K.D. Forbus and T. Hinrich, Analogy and relational representations in the companion cognitive architecture, *AI Magazine* **38**(4) (2017), 34–42.

[32] F. Ilievski, A. Oltramari, K. Ma, B. Zhang, D.L. McGuinness and P. Szekely, Dimensions of commonsense knowledge, *Knowledge-Based Systems (KBS)* (2021).

[33] F. Ilievski, J. Pujara and H. Zhang, Story generation with commonsense knowledge graphs and axioms, in: *Workshop on Commonsense Reasoning and Knowledge Bases*, 2021.

[34] J. Sun, H. Sun, T. Han and B. Zhou, Neuro-symbolic program search for autonomous driving decision module design, in: *Conference on Robot Learning*, PMLR, 2021, pp. 21–30.

[35] D. Ortiz, L.H. Gilpin and A.A. Cardenas, Semi-Automated Synthesis of Driving Rules, in: *Symposium on Vehicle Security and Privacy (VehicleSec)*, 2023.

[36] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever and P. Abbeel, Infogan: Interpretable representation learning by information maximizing generative adversarial nets, in: *Advances in Neural Information Processing Systems*, 2016, pp. 2172–2180.

[37] I. Higgins, L. Matthey, A. Pal, C. Burgess, X. Glorot, M. Botvinick, S. Mohamed and A. Lerchner, beta-vae: Learning basic visual concepts with a constrained variational framework, 2016.

[38] S. Xu, L. Mi and L.H. Gilpin, A Framework for Generating Dangerous Scenes for Testing Robustness, in: *Progress and Challenges in Building Trustworthy Embodied AI*, 2022.

[39] L. Gilpin, Reasonableness Monitors, in: *The Twenty-Third AAAI/SIGAI Doctoral Consortium at AAAI-18*, AAAI Press, New Orleans, LA, 2018.

[40] L.H. Gilpin, J.C. Macbeth and E. Florentine, Monitoring Scene Understanders with Conceptual Primitive Decomposition and Commonsense Knowledge, *Advances in Cognitive Systems* **6** (2018).

[41] L. Kagal, I. Jacobi and A. Khandelwal, Gasping for air why we need linked rules and justifications on the semantic web (2011).

[42] B. Balaji, S. Mallya, S. Genc, S. Gupta, L. Dirac, V. Khare, G. Roy, T. Sun, Y. Tao, B. Townsend et al., Deepracer: Educational autonomous racing platform for experimentation with sim2real reinforcement learning, *arXiv preprint arXiv:1911.01562* (2019).

[43] G. Singh, S. Akrigg, M. Di Maio, V. Fontana, R.J. Alitappeh, S. Khan, S. Saha, K. Jeddisaravi, F. Yousefi, J. Culley et al., Road: The road event awareness dataset for autonomous driving, *IEEE transactions on pattern analysis and machine intelligence* **45**(1) (2022), 1036–1054.

[44] E. Giunchiglia, M.C. Stoian, S. Khan, F. Cuzzolin and T. Lukasiewicz, ROAD-R: The autonomous driving dataset with logical requirements, *Machine Learning* (2023), 1–31.

[45] A.P. Shah, J.-B. Lamare, T. Nguyen-Anh and A. Hauptmann, CADP: A novel dataset for CCTV traffic camera based accident analysis, in: *2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, IEEE, 2018, pp. 1–9.

[46] K.K. Santhosh, D.P. Dogra and P.P. Roy, Anomaly detection in road traffic using visual surveillance: A survey, *ACM Computing Surveys (CSUR)* **53**(6) (2020), 1–26.

[47] L.G. Roberts, Machine perception of three-dimensional solids, PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, 1963.

[48] S. Ullman, Aligning pictorial descriptions: an approach to object recognition, *Cognition* **32**(3) (1989), 193–254.

[49] J.E. Laird, C. Lebiere and P.S. Rosenbloom, A Standard Model of the Mind: Toward a Common Computational Framework Across Artificial Intelligence, Cognitive Science, Neuroscience, and Robotics., *AI Magazine* **38**(4) (2017).

[50] J. Geyer, Y. Kassahun, M. Mahmudi, X. Ricou, R. Durgesh, A.S. Chung, L. Hauswald, V.H. Pham, M. Mühlegg, S. Dorn et al., A2d2: Audi autonomous driving dataset, *arXiv preprint arXiv:2004.06320* (2020).

[51] P. Xiao, Z. Shao, S. Hao, Z. Zhang, X. Chai, J. Jiao, Z. Li, J. Wu, K. Sun, K. Jiang et al., Pandaset: Advanced sensor suite dataset for autonomous driving, in: *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, IEEE, 2021, pp. 3095–3101.

[52] S.N. Chowdhury, R. Wickramarachchi, M.H. Gad-Elrab, D. Stepanova and C.A. Henson, Towards Leveraging Commonsense Knowledge for Autonomous Driving., in: *ISWC (Posters/Demos/Industry)*, 2021.

[53] J. Krause, M. Stark, J. Deng and L. Fei-Fei, 3d object representations for fine-grained categorization, in: *Proceedings of the IEEE international conference on computer vision workshops*, 2013, pp. 554–561.

[54] Y. Lou, Y. Bai, J. Liu, S. Wang and L. Duan, Veri-wild: A large dataset and a new method for vehicle re-identification in the wild, in: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 3235–3243.

[55] A. Robicquet, A. Sadeghian, A. Alahi and S. Savarese, Learning social etiquette: Human trajectory understanding in crowded scenes, in: *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VIII 14*, Springer, 2016, pp. 549–565.

[56] K. Eykholt, I. Evtimov, E. Fernandes, B. Li, A. Rahmati, C. Xiao, A. Prakash, T. Kohno and D. Song, Robust physical-world attacks on deep learning models, *arXiv preprint arXiv:1707.08945* (2017).

[57] A. Amos-Binks and L.H. Gilpin, Anticipatory Thinking Assessment:Stress Test Using Synthetic Data, Advances in Cognitive Systems, 2022.

[58] J. De Kleer and B.C. Williams, Diagnosing multiple faults, *Artificial intelligence* **32**(1) (1987), 97–130.

[59] J. De Kleer, Causal and Teleological Reasoning in Circuit Recognition., Technical Report, MASSACHUSETTS INST OF TECH CAMBRIDGE ARTIFICIAL INTELLIGENCE LAB, 1979.

[60] W.E. Wong, R. Gao, Y. Li, R. Abreu and F. Wotawa, A survey on software fault localization, *IEEE Transactions on Software Engineering* **42**(8) (2016), 707–740.

[61] J. Kim, B.-j. Park and J. Kim, Empirical Analysis of Autonomous Vehicle's LiDAR Detection Performance Degradation for Actual Road Driving in Rain and Fog, *Sensors* **23**(6) (2023), 2972.

[62] Y. Jeong, S. Son and B. Lee, The lightweight autonomous vehicle self-diagnosis (LAVS) using machine learning based on sensors and multi-protocol IoT gateway, *Sensors* **19**(11) (2019), 2534.

[63] H. Min, Y. Fang, X. Wu, X. Lei, S. Chen, R. Teixeira, B. Zhu, X. Zhao and Z. Xu, A fault diagnosis framework for autonomous vehicles with sensor self-diagnosis, *Expert Systems with Applications* **224** (2023), 120002. doi:https://doi.org/10.1016/j.eswa.2023.120002. https://www.sciencedirect.com/science/article/pii/S0957417423005043.
[64] J. Allen, H. Kautz, R. Pelavin and J. Tenenberg, *Reasoning about plans*, Morgan Kaufmann, 2014.
[65] J. de Kleer, M. Klenk and A. Feldman, Diagnosing Alternative Facts, in: *28th International Workshop on Principles of Diagnosis (DX'17)*, M. Zanella, I. Pill and A. Cimatti, eds, Kalpa Publications in Computing, Vol. 4, EasyChair, 2018, pp. 159–168. ISSN 2515-1762. doi:10.29007/fkwg. https://easychair.org/publications/paper/rnKw.
[66] J. Kim, A. Rohrbach, T. Darrell, J. Canny and Z. Akata, Textual Explanations for Self-Driving Vehicles, *Proceedings of the European Conference on Computer Vision (ECCV)* (2018).
[67] R. Wickramarachchi, C. Henson and A. Sheth, An evaluation of knowledge graph embeddings for autonomous driving data: Experience and practice, *arXiv preprint arXiv:2003.00344* (2020).
[68] R. Muppalla, S. Lalithsena, T. Banerjee and A. Sheth, A knowledge graph framework for detecting traffic events using stationary cameras, in: *Proceedings of the 2017 ACM on Web Science Conference*, 2017, pp. 431–436.
[69] L. Halilaj, I. Dindorkar, J. Lüttin and S. Rothermel, A knowledge graph-based approach for situation comprehension in driving scenarios, in: *European Semantic Web Conference*, Springer, 2021, pp. 699–716.
[70] D. Khashabi, Y. Kordi and H. Hajishirzi, Unifiedqa-v2: Stronger generalization via broader cross-format training, *arXiv preprint arXiv:2202.12359* (2022).
[71] J. Zhang, F. Ilievski, K. Ma, J. Francis and A. Oltramari, A Study of Zero-shot Adaptation with Commonsense Knowledge, *Automated Knowledge Base Construction (AKBC)* (2022).
[72] P. Wang, J. Zamora, J. Liu, F. Ilievski, M. Chen and X. Ren, Contextualized Scene Imagination for Generative Commonsense Reasoning, *ICLR* (2022).
[73] A. Oltramari, J. Francis, C. Henson, K. Ma and R. Wickramarachchi, Neuro-symbolic architectures for context understanding, *arXiv preprint arXiv:2003.04707* (2020).
[74] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh and D. Batra, Grad-cam: Visual explanations from deep networks via gradient-based localization, *See https://arxiv. org/abs/1610.02391 v3* **7**(8) (2016).
[75] L.H. Gilpin, V. Penubarthi and L. Kagal, Explaining multimodal errors in autonomous vehicles, in: *2021 IEEE 8th International Conference on Data Science and Advanced Analytics (DSAA)*, IEEE, 2021, pp. 1–10.