Tactical Safety Reasoning. A Case for Autonomous Vehicles.

Alexandru Constantin Serban*†, Erik Poll*, Joost Visser*†
*Radboud University, Nijmegen, The Netherlands,
†Software Improvement Group, Amsterdam, The Netherlands,
Email: *{a.serban, erikpoll}@cs.ru.nl, †{a.serban, j.visser}@sig.eu

***Abstract*—Self driving cars have recently attracted academia and industry interest. As planning algorithms become responsible for critical decisions, many questions concerning traffic safety arise. An increased automation level demands proportional impact on safety requirements, currently governed by the ISO 26262 standard. However, ISO 26262 sees safety as a functional property of a system and fails to cover emergent concerns related to autonomous decisions. In order to fill this gap we propose the field of *tactical safety*, which extends safety analysis to planning and execution of driving maneuvers, response to traffic events or autonomous system failures. It is meant to complement, not to replace functional safety properties of a system and allows the analysis of autonomous agents from a safe behavior point of view. We draw the requirements for tactical safety from an automotive standard which defines functional elements for advanced driving automation systems.**

***Index Terms*—Autonomous vehicles, Vehicle safety, Advanced driver assistance.**

I. Introduction

Autonomous driving is the leading innovation factor in today's automotive industry. New software driven companies compete with vehicle manufacturers to raise the level of driving automation. The process is often regarded as adding a layer of cognitive intelligence on top of basic vehicle platforms [1]. In this sense, software is the main innovation driver, as traditional mechanic and hardware components become a commodity [2].

Recent innovation trends promise to increase passenger's safety and traffic efficiency by minimizing human involvement and error. The transfer of total control from humans to machines is classified by the *Society of Automotive Engineers* (SAE) as a stepwise process on a scale from 0 to 5, where 0 involves no automation and 5 means full-time performance by an automated driving system of all driving aspects, under all roadway and environmental conditions [3].

As the amount of software grows, there is a need to use more advanced software engineering methods and tools to handle its complexity, size and criticality [4]. An increased automation level demands proportional impact on safety requirements, governed by the mandatory compliance to ISO 26262 [5]. However, current safety policies fail to define and represent new concerns related to *autonomous decisions*. Few researchers have exposed the challenges autonomous vehicles face as we transfer more and more control from humans to machines [6, 7, 8].

This research was founded by NWO through the i-CAVE project.

Nonetheless, it is not yet known how to reason about safe behavior of autonomous vehicles. To this end, we introduce *tactical safety*. Moreover, we seek to uncover tactical safety requirements from the SAE J3016 standard [3] and position them in a research context. SAE J3016 defines multiple levels of driving automation and includes functional definitions for each level.

In the following section we provide background information about the ISO 26262 automotive safety standard and the SAE J3016 classification of driving automation. The third section outlines a definition for tactical safety. Later, in Section IV, we follow the SAE classification in order to discover tactical safety requirements and challenges. These spread over a number of research fields discussed in Section V. We conclude and present future research in Section VI.

II. Background

Safety concerns, as stipulated by the compliance to ISO 26262, consist of the identification and management of *hazards*. In general, hazards are caused by failures. In contrast to failures in physical devices, which can occur randomly, software failures are due to design faults [9]. Therefore, in software systems, safety is achieved by avoiding or protecting against such faults.

The ISO 26262 standard provides an automotive-specific risk management framework for determining risk classes: *Automotive Safety Integrity Level* (ASIL). The ASIL definition replaces the concept of *Likelihood* from the safety risk definition (equation 1) [10] with *Controllability* (equation 2), thus enforcing how well a potential failure can be managed inside the system and not how likely it is to occur.

$$Risk = Severity \times (Exposure \times Likelihood) \quad (1)$$

$$Risk = Severity \times (Exposure \times Controllability) \quad (2)$$

ISO 26262 sees safety as a *functional property* of a system and enforces safe operation in response to inputs, hardware failures or environmental changes.

The SAE J3016 classification of driving automation for on-road vehicles, shown in Figure 1, is meant to clarify the role of the human driver, if any, during vehicle operation. It uses as first discriminant the environmental monitoring agent. In the case of partial or no automation (levels 0-2), a human driver is responsible to perceive the environment, while for higher degrees of automation (levels 3-5), the vehicle is held responsible for this task.

The more significant criteria for this paper is the responsibility for fall-back performance mechanisms. Intelligent driving systems (level 4 and 5) take full responsibility for traffic safety and fault management, while less automated vehicles (levels 1-3) require a human driver to take control in case of any fault.

Level	Name	Narrative definition	DDT		DDT fallback	ODD
			Sustained lateral and longitudinal vehicle motion control	OEDR		
Driver performs part or all of the *DDT*						
0	No Driving Automation	The performance by the *driver* of the entire *DDT*, even when enhanced by *active safety systems*.	*Driver*	*Driver*	*Driver*	n/a
1	Driver Assistance	The *sustained* and *ODD*-specific execution by a *driving automation system* of either the *lateral* or the *longitudinal vehicle motion control* subtask of the DDT (but not both simultaneously) with the expectation that the *driver* performs the remainder of the *DDT*.	*Driver* and *System*	*Driver*	*Driver*	Limited
2	Partial Driving Automation	The *sustained* and *ODD*-specific execution by a *driving automation system* of both the *lateral* and *longitudinal vehicle motion control* subtasks of the *DDT* with the expectation that the *driver* completes the *OEDR* subtask and *supervises* the *driving automation system*.	***System***	*Driver*	*Driver*	Limited
ADS ("*System*") performs the entire *DDT* (while engaged)						
3	Conditional Driving Automation	The *sustained* and *ODD*-specific performance by an *ADS* of the entire DDT with the expectation that the *DDT fallback-ready user* is *receptive* to *ADS*-issued *requests to intervene*, as well as to *DDT performance-relevant system failures* in other *vehicle* systems, and will respond appropriately.	*System*	***System***	*Fallback-ready user (becomes the driver during fallback)*	Limited
4	High Driving Automation	The *sustained* and *ODD*-specific performance by an *ADS* of the entire *DDT* and *DDT fallback* without any expectation that a *user* will respond to a *request to intervene*.	*System*	*System*	***System***	Limited
5	Full Driving Automation	The *sustained* and unconditional (i.e., not *ODD*-specific) performance by an *ADS* of the entire *DDT* and *DDT fallback* without any expectation that a *user* will respond to a *request to intervene*.	*System*	*System*	*System*	**Unlimited**

Fig. 1: SAE J3016 levels of driving automation [3].

According to SAE:

- if the driving automation system performs the entire *Dynamic Driving Task* (DDT), but a DDT *fall-back ready user* is expected to take over when a system failure occurs, then the division of roles corresponds to the 3rd level.
- if a driving automation system can perform the entire DDT and all fall-back mechanisms in limited or unlimited *Operational Design Domain* (ODD)s, then the division of roles corresponds to levels 4 and 5.

Moreover, the SAE J3016 definition for DDT outlines a list of functional components that need to be automated in order to reach level 3:

- operational functions - basic vehicle control,
- tactical functions - planning and execution for event/object avoidance and expedited route following,
- strategic functions - destination and general route planning.

The ISO 26262 standard provides a way to manage functional safety throughout the system, spanning all automation levels. However, up to this moment, few researchers have addressed the problem of safe autonomous behavior. This corresponds to tactical functions of the DDT and the DDT fall-back mechanisms. Strategic functions have no impact on traffic safety because they only handle destination and general route planning. This impacts the general travel time, but not the traffic participant's safety.

III. Tactical Safety

ISO 26262 sees safety as a functional property of a system and divides safety responsibilities between (a) drivers and (b) vehicle manufacturers. The former (a) assumes blame in case of a crash due to a human driving error, while the latter (b) is liable in case of a system failure (e.g. braking system stopped working).

As the level of automation increases, the responsibility for driving errors moves from drivers (a) to vehicles and vehicle manufacturers (b). This means that a vehicle becomes responsible for driving decisions. While all functional components that implement driving decisions have to operate safely, in conformance to ISO 26262, there is no framework to describe and reason about *safe, autonomous, decisions*. In this context we define tactical safety as a branch of autonomous vehicle safety concerned with the *safe planning and execution of driving maneuvers and response to traffic events or DDT fails*. Tactical safety is meant to complement functional safety and not to replace it.

Safety concerns of autonomous vehicles can be classified as:

- functional safety concerns - safe operation of a system in response to inputs, hardware failures or environmental changes and
- tactical safety concerns - safe planning and execution of driving maneuvers and response to driving events or DDT faults.

IV. Functional and Tactical Safety requirements in SAE J3016 context

Our definition of tactical safety covers safe planning and execution of driving maneuvers and response to traffic events when reaching a destination (1) and in response to DDT fails (2). Starting with level 0, vehicle manufacturers can deploy safety features that respond to traffic events such as emergency brake or lane departure warning. However, these are not fully responsible for either (1) or (2). In literature, such components are called *reactive safety components* and operate with a high frequency to provide minimal response to traffic. Nevertheless, they are not covered by tactical safety.

Starting with SAE level 3, a vehicle is fully responsible for (1), in the presence of a human driver, which takes responsibility for (2). A definition for (1) is not specified in the SAE J3016. However, it is implicit that no action of a vehicle can lead to a crash. Moreover, a vehicle should avoid or minimize a crash or casualties in response to other traffic participants or events. Safety components that implement (1)

are known in literature as *executive safety components* and operate at lower frequencies, constrained by the processing time. However, they are able to generate and execute complex trajectories in response to traffic events.

Figure 2 illustrates traffic situations which call for distinct safety behavior. In the first case, 2a, the blue vehicle can use a reactive component such as emergency braking to avoid a crash. In the second case, 2b, the blue vehicle can either use a reactive component or an executive component which plans and executes an over-take. However, in case 2c, the only way to avoid a crash is by choosing an avoidance trajectory, through an executive component. Finding a good balance between the two classes of components is a challenge automotive manufacturers face as more control and responsibility is transfered to vehicles.

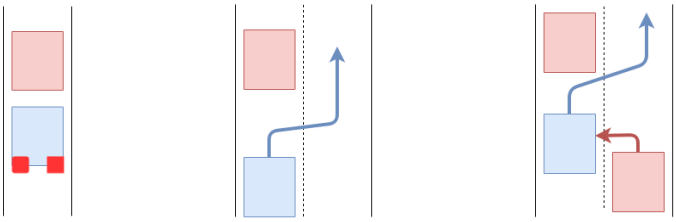

Fig. 2: Traffic situation examples.

An exhaustive, formal, definition of all traffic events and vehicle behavior is not feasible. Therefore, starting with level 3, a vehicle must *reason* about safety when planning driving maneuvers or in response to traffic events. Through reasoning, a series of future states are generated and implemented in order to satisfy an *objective*. Following the assumptions made earlier, the safety objective is to plan no action that leads to a crash and avoid/minimize the risk of a crash/casualties in response to traffic events (e.g. 2c). Safety objectives have continuous properties. A vehicle must plan to avoid or generate a crash in far future states, even though for immediate states the safety objectives are satisfied.

Given the vast ecosystem of vehicle manufacturers, it is a challenge to standardize both behavior and safety objectives. Moreover, vehicle behavior must be verified and validated before release. Future tactical safety standards for level 3 and above must define hand crafted rules or driving policies that can divide safety responsibility and assure liability. Moreover, formal tools for verification and validation are needed. As for the moment, no standardized set of requirements is articulated for safe vehicle behavior.

Even though, at level 3, an active driver assumes DDT fall-back responsibility (2), the decision to transfer the control from a vehicle to a user belongs to the automated system. This raises both functional and tactical safety challenges. At first, from a functional perspective, the system must identify a hazard that prevents normal operation. Secondly, the time needed to transfer the control from an automated driving system to a fall-back ready user has a big impact on traffic safety. A complete definition for this interval includes the time needed to detect a DDT fall-back condition, the time needed to transfer the control and the driver's response time. Moreover, the user needs time to assess the situation and gather enough data to take a decision. This is often regarded as an irony of automation [11].

Consider an automated highway driving system capable to maintain a constant speed, its lane boundaries and a distance from the vehicle in front. While engaged, this system requires a fall-back ready user. Traveling at 100 km/h, the vehicle faces a road segment which poses problems for the automated system (or the failure of a sensor) and requests the driver to take control. Given the speed and the average visual response time of a human (400 ms [12]), the vehicle will travel $\approx 11m$ until a fall-back ready user will answer to a visual signal to intervene (in theory). If we assume the user has to brake and it starts breaking in $\approx 1s$ (400ms to visually respond and 600ms to move the leg and push the breaking pedal), the vehicle will travel $\approx 28m$ before starting to break. In such cases, if an accident occurs, the blame/failure responsibility must be clearly divided between the driver and the system.

Starting with level 3 environmental monitoring is an intrinsic component of tactical safety, as an automated vehicle is fully responsible for this task. Environmental monitoring involves the acquisition and processing of huge amounts of data by advanced algorithms. Some of which are difficult to grasp. For example, neural networks image recognition algorithms are known to be error prone and difficult to understand [13]. The use of such algorithms raises serious challenges for traffic safety. Recent research shows that a change in only one pixel of an image can cause miss-classification errors [14, 15, 16]. Such perturbations may occur from malfunctioning sensors, sensor wear or malicious attackers. In order to ensure safe environmental monitoring, transparent, resilient to errors or formally provable algorithms must be used to interpret data coming from sensors. Sensor wear of malfunctioning is considered a hazard and belongs to functional safety requirements.

Only beginning with level 4 a vehicle is responsible for both (1) and (2), constrained by operational domain limits. For example, a vehicle can operate fully autonomously in limited urban areas and clear weather. Following the SAE J3016 definition of DDT fall-back, an automated driving system is responsible to bring the vehicle to a *minimal risk condition* and reduce the risk of a crash, when a fall-back intervention is needed. This extends the reasoning requirements from level 3 with DDT fall-back reasoning. The latter is meant to generate a minimum risk condition and execute necessary steps in order to reach it, while minimizing the risk of a crash. The SAE definition for DDT fall-back holds when a crash can be avoided, however, traffic situations are known to be much more complex. In cases where a minimal risk condition can not be reached or a crash can not be avoided, a vehicle must still perform maneuvers which impact general traffic safety. Objectives similar to ones introduced for level 3 vehicles must be defined for DDT fall-back reasoning.

Level 4 vehicles have limited operational domains. For example, a vehicle can operate in autonomous mode in a limited urban area, in daylight conditions. In contrast, level 5 vehicles have no restrictions. They must drive safely *under all roadway and environmental conditions* in which a human can drive. This requirement adds two dimensions to the safety reasoning concerns, expressed earlier:

- road constraints - the vehicle must be prepared to operate safety or reach a minimum risk state in all known and unknown road conditions (e.g. when a vehicle enters a road segment for which no mapping information is available).
- environmental constraints - the environmental constraints are strongly related to weather conditions, but not limited. While level 4 vehicles are bounded to limited domains, a level 5 vehicle must operate in (all) evolving weather and traffic conditions. For example, in conditions of sudden snow or unforeseen traffic conditions such as special cargoes.

Both sets of constraints impact tactical safety requirements. Where a level 4 vehicle facing an operational domain outside its bounds could cease operation, a level 5 vehicle must continue to operate safely. The impact targets both hardware (enhanced sensors are needed to satisfy all environmental and road conditions) and software components through extended reasoning capacity. For example, in traffic situation 2b and heavy rain conditions, the vehicle might decide to break instead of an over-take, because of low confidence in the way it perceives the environment.

Tactical safety requirements evolve from level 4 to level 5 through enhanced context awareness. The reasoning process for both (1) and (2) must include contextual information and must perform under uncertainty. For example, if the vehicle faces a road section for which no mapping information is available, it must continue operation towards destination or a minimal risk condition. These requirements directly impact functional safety concerns, as an enhanced ability to perceive the environment requires new functional components and sensor fusion techniques.

We summarize the tactical and functional safety requirements in Figure 3. The table follows an incremental structure, similar to the SAE J3016 classification of driving automation. All requirements from a lower level of automation are mandatory for higher levels.

SAE Level	Functional Requirements	Tactical requirements
3	Runtime hazard identification & mitigation	Error resilient algorithms for environmental monitoring; Standard safety objectives; Methods to prove correct safety reasoning in limited ODDs; Decision to transfer control; Standard & provable transfer time;
4	Runtime hazard identification & mitigation	DDT fall-back reasoning; Standard DDT fall-back objectives; Methods to prove correct DDT fall-back reasoning;
5	Runtime hazard identification & mitigation	Enhanced context awareness; Advanced methods to prove correct safety reasoning in all contexts;

Fig. 3: Safety requirements given the SAE classification of driving automation for on-road vehicles.

V. Discussion and related work

The early introduction of ISO 26262 allowed the development of tools and methods that support standard compliance [17, 18, 19]. Nevertheless, functional safety has a tangible character and relies on mechanisms for error identification and redundancy assurance. Recent developments in level 4 vehicles [20] show that full system replication is a reliable method to deal with functional safety concerns of autonomous vehicles.

Tactical safety, however, is less tangible. While it is not a requirement for machine learning algorithms to handle behavior planning or environmental monitoring, they appear as the first industrial and research choice. Recent announcements for level 4 vehicles [20], releases of driving simulators where algorithms can learn to drive [21, 22, 23] or the efficiency of such algorithms on complex tasks prove their precedence. A safe learning process recently raised community attention. Amodei et al. [24] are the first authors to compile a catalog of safety problems learning algorithms face. While the work is not automotive specific, it reveals important information about an incipient research field. Leike et al. [25] developed a simulation environment where some of the problems introduced in [24] can be mitigated and tested.

Shalev et al. [26] are the firsts to introduce a formalism for safe and scalable self driving vehicles, through the conceptualization of *blame*. In their model, a vehicle must learn to avoid blame when executing driving maneuvers, thus minimizing *accident responsibility*. Given the example in 2c, a vehicle is not responsible for the accident, thus it can take no action. Earlier, Shalev et al. [27] introduced a multi-agent learning framework for autonomous driving, arguing that in order to achieve safe autonomous driving, autonomous vehicles must learn to cooperate and work together.

Safety of artificial intelligent systems receives increasing attention because of their potential effects in safety-critical systems such as autonomous vehicles [28]. The trouble with image classification tasks, which play a crucial role in perception modules of autonomous vehicles, is that they do not have a formal specification [29]. Recent results [29, 30, 31] pave the road to formal verification of neural networks and increase their resilience to errors or attacks. However, the performance and scalability of such methods is still to evolve.

Concerned with functional safety of autonomous vehicles, Gleirscher et al. [32, 33] introduced a formalism which allows autonomous vehicles to define and reach safe states. Through this formalism, an intelligent algorithm can reason about the impact of a hazard on the vehicle planning ability and how a safe state can be reached. Moreover, the authors introduced a framework for the analysis and design of high level controllers capable of run-time hazard identification and mitigation [34]. Their work targets the behavior of an automated vehicle in

case an of an internal hazard (e.g. loss of sensors).

Despite interest in safe behavior planning, the solution domain lacks maturity. Works such as [26] open the road for safe, autonomous vehicles. We argue that, similar to functional safety, governed by the ISO 26262 standard, the reasoning framework for safe, autonomous driving, must strive for standardization. Driving policies must determine if blame, casualties or property damage are good constraints for planning algorithms.

VI. Conclusions and future research

We have outlined an exploration of safety concerns of future autonomous vehicles. We start by introducing the current automotive safety standard, ISO 26262, and the requirements for automated vehicles, as specified in the SAE J3016 standard. ISO 26262 sees safety as a functional property of a system and fails to cover emergent concerns related to autonomous decisions. In order to fill this gap, we propose the field of tactical safety, which covers safe planning and execution of driving maneuvers, response to traffic events or autonomous system failures. It is meant to complement and not replace the functional properties of a system. Moreover, it allows the analysis of autonomous agents from a safety behavior point of view.

We use the SAE J3016 standard to develop a set of tactical safety requirements and future challenges, presented in Figure 3. Few researchers have addressed the problem of safe behavior planning. While not standard, learning algorithms are the industry adopted method for trajectory planning. With little literature support, their impact on traffic safety is often overlooked.

This study is the first step towards a safety analysis of autonomous vehicles behavior. For future work we propose the development of a full reasoning framework around tactical safety that can resemble the ISO 26262 body for functional safety. In order to verify and validate tactical safety of future intelligent vehicles, formal tools and methods must be developed. While the final requirements are in the hands of international standard-setting bodies, we are interested in the development, verification and validation of safe behavior for autonomous vehicles.

References

[1] S. Behere and M. Törngren, "A functional reference architecture for autonomous driving," *Information and Software Technology*, vol. 73, pp. 136–150, 2016.

[2] M. Broy, "Challenges in automotive software engineering," in *Proceedings of the 28th international conference on Software engineering*, pp. 33–42, ACM, 2006.

[3] "Society of automotive engineers, taxonomy and definitions for terms related to on-road motor vehicle automated driving systems," in *SAE Standard J3016*, 2014.

[4] M. Staron, "Software complexity metrics in general and in the context of ISO 26262 software verification requirements," *Scandinavian Conference on Systems Safety*, 2016.

[5] "ISO 26262:2011 road vehicles - functional safety," ISO, 2011.

[6] J.-F. Bonnefon, A. Shariff, and I. Rahwan, "The social dilemma of autonomous vehicles," *Science*, vol. 352, no. 6293, pp. 1573–1576, 2016.

[7] G. E. Marchant and R. A. Lindor, "The coming collision between autonomous vehicles and the liability system," *Santa Clara L. Rev.*, vol. 52, p. 1321, 2012.

[8] F. M. Favarò, N. Nader, S. O. Eurich, M. Tripp, and N. Varadaraju, "Examining accident reports involving autonomous vehicles in california," *PLoS one*, vol. 12, no. 9, p. e0184952, 2017.

[9] W. Wu and T. Kelly, "Safety tactics for software architecture design," in *Computer Software and Applications Conference, 2004. COMPSAC 2004. Proceedings of the 28th Annual International*, pp. 368–375, IEEE, 2004.

[10] G. F. Kinney, "Practical risk analysis for safety management," tech. rep., China Lake, CA: Naval Weapons Center, 1976.

[11] L. Bainbridge, "Ironies of automation," *Automatica*, vol. 19, no. 6, pp. 775–779, 1983.

[12] S. Thorpe, D. Fize, C. Marlot, *et al.*, "Speed of processing in the human visual system," *nature*, vol. 381, no. 6582, pp. 520–522, 1996.

[13] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, "Understanding deep learning requires rethinking generalization," *arXiv preprint arXiv:1611.03530*, 2016.

[14] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," *arXiv preprint arXiv:1412.6572*, 2014.

[15] N. Papernot, P. McDaniel, I. Goodfellow, S. Jha, Z. B. Celik, and A. Swami, "Practical black-box attacks against deep learning systems using adversarial examples," *arXiv preprint*, 2016.

[16] A. Kurakin, I. Goodfellow, and S. Bengio, "Adversarial examples in the physical world," *arXiv preprint arXiv:1607.02533*, 2016.

[17] R. Rana, M. Staron, C. Berger, J. Hansson, M. Nilsson, and F. Törner, "Early verification and validation according to ISO 26262 by combining fault injection and mutation testing," in *International Conference on Software Technologies*, pp. 164–179, Springer, 2013.

[18] M. Conrad, P. Munier, and F. Rauch, "Qualifying software tools according to ISO 26262.," in *MBEES*, pp. 117–128, 2010.

[19] P. Rempel, P. Mäder, T. Kuschke, and J. Cleland-Huang, "Mind the gap: assessing the conformance of software traceability to relevant guidelines," in *Proceedings of the 36th International Conference on Software Engineering*, pp. 943–954, ACM, 2014.

[20] W. LLC., "On the road to fully self-driving." https://waymo.com/safetyreport/, 2017. [Online; accessed 01-10-2017].

[21] "Carla, open-source simulator for autonomous driving research." https://carla.org. [Online; accessed 05-12-2017].

[22] "Apolo." http://apollo.auto. [Online; accessed 05-12-2017].

[23] "Cognata, deep learning autonomous simulator." http://www.cognata.com. [Online; accessed 05-12-2017].

[24] D. Amodei, C. Olah, J. Steinhardt, P. Christiano, J. Schulman, and D. Mané, "Concrete problems in AI safety," *arXiv preprint arXiv:1606.06565*, 2016.

[25] J. Leike, M. Martic, V. Krakovna, P. A. Ortega, T. Everitt, A. Lefrancq, L. Orseau, and S. Legg, "Ai safety gridworlds," *arXiv preprint arXiv:1711.09883*, 2017.

[26] S. Shalev-Shwartz, S. Shammah, and A. Shashua, "On a formal model of safe and scalable self-driving cars," *arXiv preprint arXiv:1708.06374*, 2017.

[27] S. Shalev-Shwartz, S. Shammah, and A. Shashua, "Safe, multi-agent, reinforcement learning for autonomous driving," *arXiv preprint arXiv:1610.03295*, 2016.

[28] S. A. Seshia, D. Sadigh, and S. S. Sastry, "Towards verified artificial intelligence," *arXiv preprint arXiv:1606.08514*, 2016.

[29] X. Huang, M. Kwiatkowska, S. Wang, and M. Wu, "Safety verification of deep neural networks," in *International Conference on Computer Aided Verification*, pp. 3–29, Springer, 2017.

[30] G. Katz, C. Barrett, D. L. Dill, K. Julian, and M. J. Kochenderfer, "Reluplex: An efficient smt solver for verifying deep neural networks," in *International Conference on Computer Aided Verification*, pp. 97–117, Springer, 2017.

[31] V. Zantedeschi, M.-I. Nicolae, and A. Rawat, "Efficient defenses against adversarial attacks," in *Proceedings of the 10th ACM Workshop on Artificial Intelligence and Security*, pp. 39–49, ACM, 2017.

[32] M. Gleirscher and S. Kugele, "Defining risk states in autonomous road vehicles," in *High Assurance Systems Engineering (HASE), 2017 IEEE 18th International Symposium on*, pp. 112–115, IEEE, 2017.

[33] M. Gleirscher and S. Kugele, "Reaching safe states in autonomous road vehicles," in *Fast abstracts at International Conference on Computer Safety, Reliability, and Security (SAFECOMP)*, 2016.

[34] M. Gleirscher and S. Kugele, "From hazard analysis to hazard mitigation planning: The automated driving case," in *NASA Formal Methods Symposium*, pp. 310–326, Springer, 2017.