A Systematic Literature Review About the Impact of Artificial Intelligence on Autonomous Vehicle Safety

Alexandre Moreira Nascimento, Lucio Flavio Vismari, Caroline Bianca Santos Tancredi Molina, Paulo Sergio Cugnasca, João Batista Camargo, Jr., *Member, IEEE*, Jorge Rady de Almeida, Jr., Rafia Inam, Elena Fersman, Maria Valeria Marquezini, and Alberto Yukinobu Hata

***Abstract*— Autonomous Vehicles (AV) are expected to bring considerable benefits to society, such as traffic optimization and accidents reduction. They rely heavily on advances in many Artificial Intelligence (AI) approaches and techniques. However, while some researchers in this field believe AI is the core element to enhance safety, others believe AI imposes new challenges to assure the safety of these new AI-based systems and applications. In this non-convergent context, this paper presents a systematic literature review to paint a clear picture of the state of the art of the literature in AI on AV safety. Based on an initial sample of 4870 retrieved papers, 59 studies were selected as the result of the selection criteria detailed in the paper. The shortlisted studies were then mapped into six categories to answer the proposed research questions. An AV system model was proposed and applied to orient the discussions about the SLR findings. As a main result, we have reinforced our preliminary observation about the necessity of considering a serious safety agenda for the future studies on AI-based AV systems.**

***Index Terms*— Autonomous vehicles, safety, artificial intelligence, machine intelligence.**

I. Introduction

ADVANCES in Artificial Intelligence (AI) are one of the key enablers of the Autonomous Vehicles (AVs) development. In fact, AVs rely on AI to interpret the environment, understand its conditions, and make driving-related decisions. Thus, it basically replicates the human driver actions when driving a vehicle. In this context, AI applied to AV has become an important research topic.

AV is a safety-critical system. When operating in an undesirable way, AV can jeopardize human lives or the environment in which it operates. It has the potential to threaten the lives of its own passengers, pedestrians and people in other vehicles, and damage other transportation system elements (e.g. other vehicles and transportation infrastructure). Therefore, it is mandatory to assure AV is safe, mainly when operating on public roads in which resources will be shared with other systems (and people).

Although safety is a mandatory characteristic to AV, and although the researchers seem to agree on the importance of AI applied to autonomous vehicles, they seem to disagree on the AIs impact on AV safety. Many researchers, in special those related to the AI community and AV manufacturers, advocate AI as one of the core elements to enhance AV safety. Their hypothesis is the automation of the driving tasks will lead to a significant reduction of the car accidents. However, other researchers, mainly in the system safety community, argue that AI can potentially jeopardize AVs safety.

This study is the first, as far as we are aware, to map and to organize the related literature and to provide a complete view of the aspects related to both visions, and to subsidize future studies. A preliminary study on the concerns about the differences between AI and system safety mindsets impacting AV safety was published in [1]. In this non-convergent context, this paper presents a systematic literature review (SLR) aiming to present a clear picture of the state of the art of the literature in AI on AVs safety.

This paper is structured into 5 sections. Section II presents details about the research methodology used. Section III presents the data analysis results from the SLR based on the proposed methodology. Section IV proposes an AV system model that is used to orient the discussions about SLR findings. Finally, Section V presents the conclusions.

II. Research Methodology

This study was performed using the systematic literature review (SLR) method. The reasons supporting the SLR use are: (1) its established tradition as a tool to understand state-of-the-art research in technology-related fields [2]; (2) it helps to understand existing studies and supports readers in identifying new directions in the research field [3]; and (3) it helps to create a foundation for advancing knowledge [4].

The protocol used (Fig. 1) was based on the tasks suggested by [5], [6] for defining the research questions, identification

Manuscript received April 17, 2019; revised August 7, 2019; accepted October 2, 2019. Date of publication November 6, 2019; date of current version November 30, 2020. This work was supported by the Research, Development and Innovation Center, Ericsson Telecommunication S.A., Brazil. The Associate Editor for this article was S. C. Wong. *(Corresponding author: Lucio Flavio Vismari.)*

A. M. Nascimento, L. F. Vismari, C. B. S. T. Molina, P. S. Cugnasca, J. B. Camargo, Jr., and J. R. de Almeida, Jr., are with the Safety Analysis Group, Polytechnic School, University of São Paulo, São Paulo 05508-010, Brazil (e-mail: alexandremoreiranascimento@alumini.mit.edu; lucio.vismari@usp.br).

R. Inam and E. Fersman are with Ericsson Research, Ericsson AB, 164 53 Stockholm, Sweden (e-mail: rafia.inam@ericsson.com).

M. V. Marquezini and A. Y. Hata are with Ericsson Telecommunication S.A., Indaiatuba 13337-300, Brazil (e-mail: maria.marquezini@ericsson.com).

Digital Object Identifier 10.1109/TITS.2019.2949915

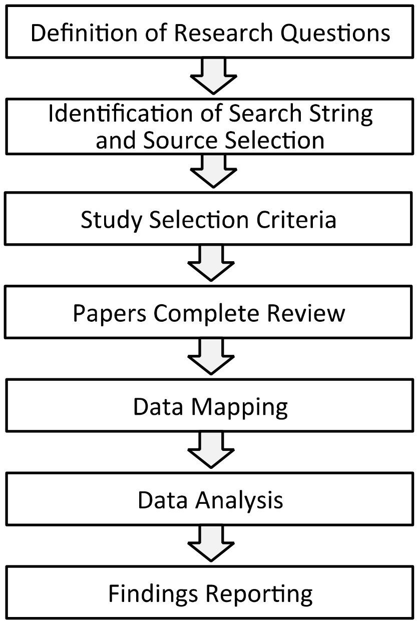

Fig. 1. Protocol used to support systematic literature review.

TABLE I
NUMBER OF PAPERS PER DATABASE

Database	#Entries
ACM	36
Engineering Village	191
ScienceDirect	81
Scopus	182
SpringerLink	3999
Wiley	329
WoS	52
Total	**4870**

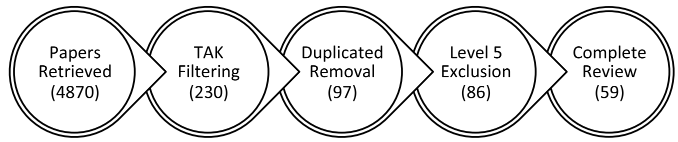

Fig. 2. Study selection process.

of search string, source selection, study selection criteria, and data mapping. Also, the protocol followed the recommendations of [4], [7]–[9] for extracting, analyzing, interpreting and reporting the literature-based findings.

A. Definition of Research Questions

The first step was to define the research questions (RQ). In order to support the research goal of presenting a clear picture of the state of the art in the literature about AI on AV safety, the following research questions were posed:

- RQ1.How do AI-based systems impact system safety?
- RQ2.Which are the topics (context domain) of the studies identified?
- RQ3.Which AI-related techniques are used on the studies?
- RQ4.Which problems do the techniques seek to address?
- RQ5.Which findings are reported by the study's authors?
- RQ6.Which future studies are suggested in these studies?

B. Identification of Search String and Source Selection

The search strategy was structured through the selection of source databases and the appropriate search terms. No date range was used, to ensure that relevant studies were covered, regardless of their publication date. A broad selection of online databases indexing scientific literature was considered: ACM, Engineering Village, ScienceDirect, Scopus, SpringerLink, Wiley and Web of Science (WoS). Please note that IEEExplore is already covered by the selected databases for this SLR study.

The search string was designed based on the synonyms of the 3 main concepts related to the investigated topics: Safety, Artificial Intelligence and Autonomous Vehicle. Many synonyms are present in the literature for the terms "artificial intelligence" and "autonomous vehicle". Therefore, an exploratory study of their most representative synonyms was performed. Then, a careful selection of synonyms was made to ensure the search process would have an appropriate coverage. As a result, the following string with Boolean operators was selected: *("safety" AND ("artificial intelligence" OR "machine intelligence" OR "machine learning") AND ("autonomous vehicle" OR "autonomous car" OR "automated vehicle" OR "automated car" OR "self-driven vehicle" OR "self-driving" OR "driverless"))*. Note that the synonyms for each one of the topics are already presented in the Boolean string previously displayed.

Different instances of the search string were created to adapt it to the distinct database search syntax rules, but the same logical value was kept. In each database, the appropriate options were selected to limit the search process to the Title- Abstract-Keyword (TAK) field set. This is an important measure to reduce the number of non-related or duplicated studies retrieved. However, it was observed that not all databases support a search limited on TAK field set, leading to an inflated number of papers found (e.g. SpringerLink). Table I shows the initial number of papers found per database.

C. Study Selection Criteria and Papers Review

The study selection process is shown in Fig. 2. Each step indicates the number of papers remaining as a sample after the corresponding step was executed. The first selection criterion applied was to ensure that only the studies with the TAK fields returning positive to the Boolean search expression would be selected. The information (metadata) available for each paper found, in the first step of the selection process, was collected by exporting the results to a spreadsheet. A spreadsheet macro was developed to analyze the TAK fields and to properly select the papers. After this check, only 230 papers remained as a sample. Using the spreadsheet Remove Duplicate tools, the duplicated entries were removed. The 97 remaining papers composed the selected sample.

As a reasonable number of papers (97) was found [10], book chapters, editorials, notes or reports were excluded - level 5 exclusion [10] - and 86 papers remained. The abstracts, titles and keywords of the remaining 86 peer reviewed papers were scrutinized to check their fitness with the goals of this research. After a careful examination (sometimes a full-paper skimming was necessary), 27 papers were considered not related to this research and were excluded from the sample of the literature mapping. Finally, a sample of 59 papers was considered for this study.

There was a considerable drop in the number of studies, from the initial 4870 to the final 59 papers selected. It occurred for different reasons, such as: misuse of the terminology; correct use of the terminology in the context of an example within a paper that did not actually focus on the topic; or lack of restricted search in TAK fields in some databases (in our study, the SpringerLink).

D. Data Mapping

The 59 selected papers were completely reviewed and scrutinized. For each paper, the required information for answering each research question was retrieved and placed into a spreadsheet. Subsequently, aiming to support a proper data organization and its quantitative analysis, a codification process was executed.

In order to facilitate the information normalization and to guide the mapping process, 6 categories were created: CT1 (Impact), CT2 (Topics), CT3 (Techniques), CT4 (Problem), CT5 (Findings), and CT6 (Future Studies). Each one of these categories corresponded to a unique research question (see Section II-A). Furthermore, each category encompassed a set of code values, which different strategies were used to create them (Table II):

- For CT1 (Impact), the code "increase" was used when the paper described AI as a factor of increasing the safety risk (negative impact on safety); and the code "decrease" was used when the paper presented AI as a factor of decreasing the safety risk (positive impact on safety).
- For CT2 (Topics), CT5 (Findings) and CT6 (Future Studies), the codifications were derived by the context domain of the study according to what was reported by their authors, as suggested by [11].
- For CT3 (Techniques) and CT4 (Problem), the codes were based on what was reported by their authors [11] and, due to the wide range of techniques, subfields and misuses of terms, the terminologies were mapped and normalized according to the literature references [12]–[17] in the field.

Finally, all the retrieved information was mapped into those codes to answer the research questions in a normalized way. The codification process was based on the agreement of researchers working in this study. More information on the codification process for each specific category is provided in Section III.

III. Data Analysis

The distribution of the studies over the years can provide an overview of the size and evolution of the field (Fig. 3).

TABLE II
DATA MAPPING STRUCTURE

RQ#	Category	Code Values	Reference
RQ1	CT1 Impact	"Increase"/ "Decrease"	Defined by the authors of present study
RQ2	CT2 Topics	Adapted from what was reported by papers authors	[11]
RQ3	CT3 Techniques	Adapted from what was reported by papers authors according to the AI field literature and normalized according to the literature references	[11], [12]–[17]
RQ4	CT4 Problem		
RQ5	CT5 Findings	Adapted from what was reported by papers authors	[11]
RQ6	CT6 Future Studies		

The left chart in Fig. 3 shows the distribution from 1987 until 2018 (April). The oldest study found dates back to 1987. No work was found for over a decade – from 1991 to 2002 – considering the adopted search criteria.

This period can be labeled as the "first winter" in this research topic as an analogy to the Artificial Intelligence "winters".[1] Only one paper a year was found over the following 3 years – from 2003 to 2005. A second short winter was found from 2006 to 2008. Only 1 paper was found in 2009 and another in 2011, while no paper was found in 2010. Finally, the combination of AI, safety and autonomous vehicles started to get more attention from the scientific community in 2012 when 5 papers were found, although no paper was found in 2013. In fact, 51 of the papers (86%) found were published from 2012 to 2018.

The right chart in Fig. 3 shows the distribution of the studies over the last decade. The year 2018 was excluded from the plot to avoid misinterpretation. Considering the results presented in Fig. 3, the field is gaining momentum based on the continuous growth in the number of published studies since 2014. The trend line built in the last decade data shows a higher angular coefficient, indicating the momentum in recent years.

Most of the papers found are from conference proceedings. In fact, 45 papers (76%) are from conferences. Only 14 papers (24%) were published by journals. Therefore, it is reasonable to expect a growth in the number of publications about this topic in journals. Besides evaluating the time distribution of papers, another important aspect is the consistency-check of the selected keywords in the papers considered. This was performed by checking the most representative keywords

[1] The Artificial Intelligence field had periods of warm enthusiasms and some periods of very low enthusiasm, with a much lower number of publications and contributions. The literature named those low enthusiasm periods as AI winters.

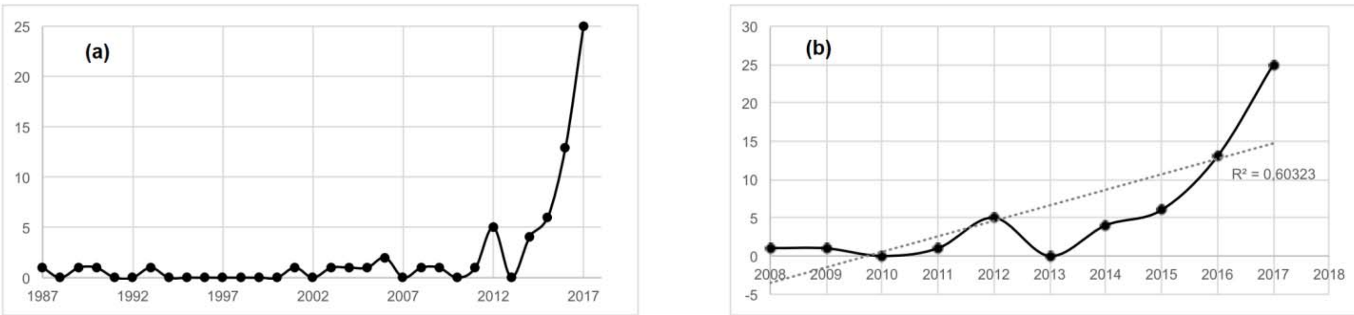

Fig. 3. Studies distribution over the years: a) depicting all studies till 2018; b) depicting studies in the last ten years.

TABLE III
KEYWORDS HITS

Concept	Keyword	#Hits	%Papers
Safety	Safety	59	100
Artificial Intelligence	Artificial intelligence	36	61
	Machine learning	27	46
	Machine intelligence	1	2
Autonomous Vehicle	Autonomous vehicle	37	63
	Automated vehicle	12	20
	Self-driving	11	19
	Autonomous car	6	10
	Driverless	2	3
	Autonomous truck	1	2

among all the synonyms of each of the three sets (previously presented) in the search string. All the keywords from the search string found on each paper TAK were accounted. As a result, the total number of hits per keyword was computed. Table III shows the number of studies with each keyword present (hits per keyword) and the percentage of the 59 sample papers with the keyword. Note the sum of the number of hits does not totalize 59. Also, the sum of the percentages for all the keywords for each distinct concept does not totalize 100%. This is because many papers have more than one synonym present, which makes it be accounted more than once.

Thus, it is possible to note the most representative keyword for each concept: safety, artificial intelligence and autonomous vehicle. In fact, a search string using only those keywords would result in 36 papers, which corresponds to 61% of the sample size of the present study. However, many other keywords used could not be ignored, since they have a considerable representativeness, such as: machine learning, automated vehicle, self-driving and autonomous car. Conversely the keyword autonomous truck surprisingly had only one hit. The following sub-sections present the results for each research question (RQ1-6).

A. AI-Based Systems Impact on Safety (RQ1)

The RQ1 was answered with the categorization of the sample studies into CT1 (Impact). Most studies consider AI a technology that increases the system safety (positive impacts on safety). So, 81% (48) of the papers were actually coded as decrease, because they argue that AI decreases the safety risks. Only 19% (11) of the studies consider AI a potential threat to the system safety.

B. Main Topics of the Studies (RQ2)

In order to answer RQ2, the sample papers were classified into the category CT2 (Topics). Studies were grouped based on their CT1 coding into two distinct sets: Increase Safety Risks and Decrease Safety Risks. Then studies were grouped by their similarities and each group was coded with a label that could encompass all its members. Table IV shows the results of this coding process.

As observed, the papers positioning AI as a factor that decreases safety risks (48 papers, 81%), they studied the subjects related to five main topics: Sensors and Perception (22 papers, 46%), Navigation and Control (11 papers, 23%), Fault Prevention (6 papers, 13%), Conceptual Model and Framework (4 papers, 8%) and Human Factor (5 papers, 10%). In turn, the papers positioning AI as a risk to system safety (11 papers, 19%) studied subjects related to three main topics: Fault Forecasting (5 papers, 45%), Ethics and Policies (4 papers, 36%) and Dependability and Trust (2 papers, 18%). The complete list of references for each code in this category can also be found in Table IV.

The main topics for each group of papers differ reasonably from each other. While the papers in the category decrease focus on important aspects to support or to enhance the vehicle autonomy, the papers in the category increase (endanger safety) focus on topics related to safety assurance.

Sensors and Perception is the topic with the largest number of studies (22). They are mostly related to computer vision and detection techniques necessary for adding the necessary capabilities to detect different aspects of the navigation environment and supporting the autonomy of the AVs, such as: general computer vision [18], Doppler sensing [19], lane detection [20], daylight detection and evaluation [21], obstacles detection [22]–[24], pedestrian detection [25], [26], pedestrian trajectory prediction [27], road detection [23], [28], [29], road junction detection [30], road terrain detection [31], traffic signal detection [32], [33], situation awareness [34], speed bump detection [24], [35], traffic light detection [36], vehicle detection [37] and virtual worlds for training detection [38].

The second largest number of papers (11) found encompasses studies related to Navigation and Control. They are

TABLE IV
IMPACT OF AI-BASED SYSTEMS ON SAFETY AND ITS MAIN TOPICS AND REFERENCES

Category	Codes	#Hits	%Papers	References
CT1 - Impact	Decrease Safety Risks (Positive Impact on Safety)	48	81	
CT2 - Topics	Sensors and Perception	22	44	[18]–[38], [46]
	Navigation and Control	11	27	[39], [40], [42]–[45], [47]–[50], [77]
	Fault Prevention	6	13	[39], [52]–[56]
	Human Factor	5	10	[61][57]–[60]
	Conceptual Model and Framework	4	8	[62]–[65]
CT1 - Impact	Increase Safety Risks (Negative Impact on Safety)	11	19	
CT2 - Topics	Fault Forecasting	5	45	[66]–[70]
	Ethics and Policies	4	36	[71]–[74]
	Dependability and Trust	2	18	[75], [76]

mostly related to techniques necessary to ensure the proper autonomous navigation and control capabilities required by AVs, such as: adaptive pre-crash control [39]; safe trajectory selection [40]; AV following another car driven by a human pilot (Trailing) [41]; safe navigation [42]; heuristic optimization algorithm for unsigned intersection crossing [43]; vehicle coordination [44]; maneuver classification [45]; learning to navigate from demonstration [46]; AV movements optimization in intersection [47]; learning and simulation of the Human Level decisions involved in driving a racing car [48]; path tracking [49]; and fuzzy-logic control approach to manage low level vehicle actuators (steering throttle and brake) [50].

Six papers with research related to Fault Prevention were found. These studies encompass researches related to the preventing the occurrence or introduction of faults [51], such as AI for security of wireless communication to ensure safety [52]; remote diagnosis, maintenance and prognosis Framework [53]; prediction of computational workload [54]; vehicle security against cyber-attack [39], [55]; and diagnosis of sensor faults [56].

Five studies on Human Factor and four on Conceptual Model and Framework were found. The studies on human factor cover important aspects to be considered in the autonomous cars engineering due to the human-in-the-loop factor, such as: safety, comfort, and stability based on the human driver perception behavior [57]; design of real time transition from assisted driving to automated driving under conditions of high probability of a collision [58]; diagnosing and predicting stress and fatigue of driver in semi-automated vehicles [59]; advances in driver-vehicle interface [60] and remote-controlled semi-AV based on IoT [61]. Considering the studies (4) proposing conceptual models and frameworks, they have a considerable diversity of focus, such as: ML and cloud-based framework proposed to address safety and reliability-related issues [62]; AV conceptual model [63]; an interdisciplinary framework to extract knowledge from the large amount of available data during driving to reduce driver's behavioral uncertainties [64]; and a proposition of an AV highway concept to improve highway driving safety [65].

Considering the group of papers positioning AI as a potential factor of decreasing the safety, the highest number of studies was related to Fault Forecasting. In other words, those papers dealt with the limitations to estimate the present number and future incidence of faults in AI-based systems, by executing activities related to evaluation, testing, verification and validation [51], such as: aspects (and limitations) related to safety validation [66]; performance and safety verification methodology [67]; test suites for AV [68]; end-to-end safety for AV design [69]; and a framework to evaluate the impacts of such a sophisticated system on traffic and the impact of continuous increase in the number of highly automated vehicles on future traffic safety and traffic flow [70].

There were four studies related to discussions about Ethics and Policies. One of the studies discussed and performed experiments on how distinct ethical frameworks adopted to make decisions about AV crashes can affect the number of lives endangered [71]. The other studies discuss the scope of AI on AV with ethical aspects [72], ethics in AV design [73], and moral values and ethical principles for autonomous machines [74]. As can be seen, those studies are quite recent since the oldest one was published in 2015.

Finally, 2 papers were found related to Dependability and Trust. Dependability is an important concept in critical systems, because it comprises attributes such as safety, security, availability, reliability and maintainability, as well as how (the mechanisms) to keep these systems attributes [51]. According to [51], trust can be defined as accepted dependability. The studies found are thus related to: safety issues [75] and current mechanisms to ensure robust operation in safety-critical situations facing the introduction of non-deterministic software [76].

C. Techniques Used (RQ3) and Problems They Seek to Address (RQ4)

Aiming to answer RQ3 and RQ4, the sample papers were classified into categories CT3 (techniques) and CT4 (problems) based on how their authors described the AI technique used in the study. Then, some terminologies used to define the codification for the categories CT3 and CT4 were adapted based on the field literature [12]–[17],when necessary.

Most reviewed papers reported the specific AI related techniques used in the research. Some reported the use of more than one technique, whereas others reported only the approach used. Some papers (14 papers, 24%) were related to general aspects of AI or ML techniques, without mentioning specific techniques used or researched [18], [59], [63]–[67], [69]–[74], [76].

All the techniques found in the reviewed papers were mapped considering the problem (CT4) that they were solving. As a result, Table IX - placed in the Appendix - lists the techniques found, the number of papers in which they were used, the main problems they were seeking to address, and the references.

As can be seen, there is a considerable number of studies (13 papers, 22%) that used techniques related to artificial neural networks. Also, there is a reasonable number of studies reporting the use of SVM (10 papers, 17%). Some studies used Fuzzy Logic (5 papers, 8%), Bayesian Artificial Intelligence (e.g. Bayesian Deep Learning, Naive Bayes Classifier-NBC, etc.) (4 papers, 7%), Hidden Markov Based Models (e.g. Continuous Hidden Markov Model-CHMM and Discrete Hidden Markov Model–DHMM) (4 papers, 7%), Estimation Filters (e.g. Kalman Filter and Particle Filters) (4 papers, 7%), Nearest-Neighbors-Based Algorithm (e.g. k-Nearest Neighbors - kNN) (4 papers, 7%), Adaptive Boosting (AdaBoost) (3 papers, 5%), Ramer-Douglas-Peucker or Ramer Douglas algorithm (3 papers, 5%), Haar-like feature detector (3 papers, 5%), Histogram of Oriented Gradient (HOG) (3 papers, 5%), Hough Transformation (3 papers, 5%), Optimization Heuristics (3 papers, 5%), Regression-Based Models (3 papers, 5%) and Principal Components Analysis (PCA) (2 papers, 3%).

Analyzing Table IX, it shows that each of the following techniques were reported, in all the reviewed papers, only once: Canny Edge Detection Algorithm, Case-based reasoning (CBR), Channel Features, Clustering Algorithm k-mean, Complex Decision Trees (CDT), Conditional Random Fields (CRFs), Distributed Random Forest (DRF), Gaussian Mixture Model (GMM), Linear Temporal Logic (LTL), Local Binary Patterns (LBP), Neuroevolution of Augmenting Topologies (NEAT), Novel Image Recognition Technique, Path Planning Algorithms (A^* and $D*$), Satisfiability Modulo Theories (SMT) Solver, and Viterbi Algorithm. Thus, there is room for new studies using techniques not yet used or under-represented by the set of papers considered.

D. Reported Findings (RQ5)

Question RQ5 is answered by CT5 (findings), based on the information about the findings reported on the sample papers. Some papers did not report specific main findings in a straightforward way because the propose frameworks or approaches had not yet been tested or the results were still incipient. Other papers described very specific findings that would require a background section to support a proper discussion. In those cases, only a higher level of abstraction of the results is presented. Finally, because of space limitation, only some specific examples are described here, while most of the results are presented grouped around the main topic of research. A complete list, oriented by the discussion presented at Section IV, can be found in [78].

The papers about topics related to Sensors and Perception presented positive and promising results with the techniques employed to address their research problems. In fact, this topic already achieved significant results with the recent developments in AI and sensor technologies. While AI had the image and pattern recognition boosted by advancements such as the new architectures of ANNs and new machine learning techniques, sensor technologies have been boosted in the last decades by the advancements in the robotics and mobile phone industries. As a result, the papers demonstrated applications of enhancements in the techniques or combination of techniques and sensors in order to recognize and to detect important elements and signals the human drivers need to handle to ensure the proper operation of a vehicle. In this context, the findings are positive for the application of ANNs to recognize turn signal [33], road environment and signals [26], [30]–[32], and pedestrian [25], [27], for example. Likewise, some papers reported SVM has been applied successfully to detect road [29], traffic light [36], and pedestrian [26].

The papers related to Navigation and Control also reported positive and promising results. As presented previously, they used diverse AI techniques to seek to address a broad range of problems. For example, a hybrid AI architecture encompassing ANN, CBR, and a hybrid Case-Based Planner (A^* and D^* motion planner) was successfully tested to tackle the pre-crash problem of intelligent control of autonomous vehicles [39], while SVM was used to support a safest path planning in a dynamic environment to avoid maneuvers too close to an obstacle [42].

Each one of the 6 papers on Fault Prevention used a distinct AI technique for the research problems. One paper presented a preliminary result [54], and another one proposed an approach but did not report results [56]. All the others papers, related to the detection of cyber-attack, presented promising positive results for the application of ANNs [39], Estimation Filters [52], and Fuzzy-Logic [55], for example. Also, preliminary positive results have been reported on the use of a regression-based model to predict the CPU patterns [54].

Two from the five papers related to the topic Human Factor, have presented preliminary positive results. One presented promising results from using a regression-based model to deal with selective attention mechanism [57], while the other presented some examples of scenarios where the use of Bayesian AI could avoid the collision when no action is taken by the human driver [58]. The other 3 papers did not present specific findings, due to their theoretical nature related to the design considerations for the driving assistance system [60] and human drivers monitoring to enhance the integration between AVs and human drivers [59], or due to their proof-of-concept nature [61].

The papers proposing conceptual models and frameworks did not present findings related to experimental results. Most of them relied on general AI/ML instead of a specific technique [63]–[65]. Also, besides the proposed approaches themselves, they focused the discussions around the issues they aimed to address, the theoretical background and future potential problems to be addressed in the field.

The last three topics (Fault Forecasting, Ethics and Policies, and Dependability and Trust) have papers more oriented to theoretical discussions and propositions around the challenges AVs are facing or will face related to safety topics, such as test and validation [66], certification [69], [76], autonomy assurance and trust when non-deterministic and adaptive algorithms are used [76] - crash assignment facing distinct ethical theories [71], for example. In this context, most of them do not present specific findings using experimental setups; instead, they envision potential future solutions for the discussed challenges. In other words, those papers try to shed an alert light on the important topics that seem to be neglected by the AV enthusiasts, trying to push the research agenda towards safety engineering mindset.

As exceptions, 3 papers presented practical applications and results. Reference [71] presented some interesting findings using a simple experimental simulated environment to test specific crash scenarios under three ethical theories. They found that understanding rational ethics is crucial for developing safe automated vehicles. The results of their experiment indicate that in specific crash scenarios, utilitarian ethics may reduce the total number of fatalities that result from automated vehicle crashes. [68] proposed an approach to describe test-cases for validating autonomous vehicles using recordings of traffic situations for creating a minimal test-suit that could help in the certification process. Considering the example presented, they show how minimalism is achieved by manually comparing the test-cases. Although it is an interesting and promising approach, there are no evidences that it could address a safety certification processes requirement when considering non-deterministic algorithms. Hence, the research was still preliminary. Finally, although [75] presents an end-to-end Bayesian Deep Learning architecture to reduce the risks of hard classifications by adopting probabilistic predictions accounting for each model, no findings from real experiments were presented.

E. Reported Future Studies (RQ6)

Question RQ6 is answered by CT6 (future studies), based on the collected information about future studies reported on the sample papers. Some papers did not suggest future studies. Other papers described intended future studies or works under development. Those are frequently small incremental changes, such as change of parameter or new test scenarios. Therefore, they are not reported here since their specificities would require a considerable background on their contents. That is out of the scope of the systematic literature review.

The studies related to Sensors and Perception propose many future studies, but mostly around improvements that would be made in the future to address some of the limitations of the presented research. Due to the space limitations, only some examples are described here. Reference [33] suggests additional research on image recognition of low contrast images and vehicle images with brake lamps. Reference [36] suggests future work on traffic lights detection under severe weather or night conditions. Reference [35] suggests more research on detecting speed bump during night time. They also suggest research on speed bumper detection when they have no pattern or marking. In addition to that, [35] suggests research to improve the recognition capabilities to distinguish zebra crossing from speed bump. Reference [29] proposed future research about road detection using road lane markers that could be detected by LIDAR, while [20] proposed more research focused on optimizing the lane detection and vehicle recognition algorithms to reduce their computational costs. Also considering the high computational costs, [26] proposed using parallel computing to increase the speed of the image recognition algorithms. Finally, according to [38], additional research is needed on using the virtual environments for testing because the authors believe their usage for training and testing intelligent systems are becoming more relevant.

Most of the studies related to Navigation and Control suggest future studies. The majority suggests extensions to the work they presented. Here, few examples are presented. The study proposing hybrid control architecture [39] suggests an extension to consider the full kinematics and dynamic limitations of the vehicle, while constantly acting to avoid collisions and unsafe driving. The paper proposing an approach using SVM to avoid maneuvers too close to an obstacle by adding a safety margin [42] proposes future re- search to extend it using a combination with the kinetic convex hulls[2] to enable the possibility of computing the solution ahead in time. According to the authors, this would help to predict the position and the width of the optimal margin. As a result, it would improve the approach by adding the ability of reduce the collision risk by preventing the AV from driving into a dangerous situation. The study using Fuzzy Logic as the main approach to control a semi-autonomous car 100-km experiment [77] proposes future research using new sensors and filtering methods for data fusion to reduce the risk on scenarios where the GPS signal is lost. Finally, the study on AVs intersection crossing [47] describes future work in which more types of vehicles and more adjacent intersections would be included in the simulations.

Most of the studies (4 of 6) related to Fault Prevention suggest future studies. Half of the studies are related to security aspects, while the other half is related to diagnosis/prognosis/prediction. The study proposing a cyber-attack detection system based on ANNs [39] suggests a future study to apply the proposed approach to a real vehicle in addition to the application of LSTM to detect online sensor attack. The study proposing the use of Particle Filter and Kalman Filter to secure connected vehicles against DoS attack [52] proposes future work to assess the proposed security scheme under many distinct scenarios, and also to execute tests in real world set-ups. The study about predicting ADAS remaining useful life for the prognosis of its safety critical components using ANNs and other techniques, such as SVM [53], proposes a considerably wide range of future studies, such as using Least Square Support Vector Machine (LS-SVM); using big data techniques to analyze the server data; studying connected vehicle prognosis; using driver, vehicle and region profile data to understand the impact on the environment and

[2]Check [42] for more information about kinetic convex hulls.

driving style impact on the system lifespan; and more studies on prognostics-enabled decision Making (PDM). Finally, the work presenting the use of regression-based methods to predict the CPU usage patterns of software tasks running on an AV [54] suggest future work on the use of some regularization methods for automatic feature selection, but also to particularly investigate the effects of under-estimating CPU utilization, and how to handle under-estimation of CPU utilization when it happens, aiming to better understand how safe over (or under) estimation of CPU utilization is in terms of reliable autonomous driving.

The studies about Ethics and Policies on AVs basically suggest more research on those topics. In the same way, most of the studies tackling human-factor-related topics do not propose future studies. As an exception, the paper proposing the application of regression-based model for the selective attention mechanism subject [57] proposed a future study to help to reveal the mechanism of rear end collision accidents.

Half (2 of 4) of the studies related to Conceptual Model and Framework do not suggest any future studies. However, implicitly, the next steps would be the deployment of those suggested approaches on experimental set-ups to collect real results. The study proposing a framework to reduce the uncertainty of a driver behavior prediction model [64] suggests more studies focusing on the resilience and sustainability of the system when deployed on a large scale in a complex system.

The papers about Fault Forecasting suggest some future research. Among them, [66] suggests more research on safety envelope mechanisms to describe a boundary within the state space of the AVs rather than trying to prove that it will always work correctly. Koopman, in another paper [69], suggest that the accepted practices must be updated to create an end-to-end design and validation process to address all the safety concerns considering cost, risk, and ethical considerations. Reference [68] proposes more work on creating automated test-cases. Reference [70] proposes more studies based on the framework they proposed to evaluate the impacts of AVs on traffic safety, specially using stochastic simulations with random number seeds to achieve a broader representative and a variety of traffic situations, as well as using the proper statistical analysis techniques to ensure the statistical validity of the results.

Finally, the 2 studies about Dependability and Trust also present some suggestions of future studies. Reference [75] asks for more research on new concrete safety evaluation metrics. Reference [76] suggests more research on understanding the dependence of the system components on AVs is needed to establish trust. They also suggest that could be achieved by investigating the many ways in which people, the system, and the environment interrelate.

IV. SLR Findings Oriented by an AV System Model

In the previous section, the state of the art in the literature about AI on AV safety was identified and investigated by means of a SLR. Six research questions oriented the literature identification, in which studies that include keywords related to safety, AI and AV were considered. The resulting studies were investigated and mapped into 6 categories: Impact (increase or decrease safety risks), Topics (sensors and perception; navigation and control; fault prevention; conceptual model and framework; human factor; etc.), Techniques (general AI/ML; ANN; SVM; etc.), Problem (AV validation; road detection; collision avoidance; etc.), Findings, and Future Studies.

These results considered the AV as a system, but its specific components and functions in an architectural point-of-view were not considered. For deepening the understanding about the state of the art of AI on AV safety it is necessary to show how the presented works are applied/fitted on AV in an architectural point-of-view. In other words, which of AV modules/components and functions are already being developed and which one could be more explored. In order to achieve this goal, it is going to be considered the AV architecture proposed in [79].

An automotive manufacturer consortium (CAMP-AVR) [79] proposed a high-level architecture considering the main system components demanded for the vehicle movement control, to be used in the deployment of future Dynamic Driving Tasks (DDT). Fig. 4 (left) illustrates the model considering a traditional vehicle (i.e. human operation with no automation deployed), and Fig. 4 (right) illustrates the introduction of some level of machine automation (hybrid) in Sensors, Controller and Actuators elements. While the diagram considering the human operation can solely be mapped to the SAE Automation Level 0 (no automation), the hybrid one encompassing machine automation with human-in-the-loop can be mapped to the SAE Automation Levels 1 to 4 (semi-autonomous) [79].

In this context, a modified version of the semi-autonomous model is proposed here (Fig. 5) including the system boundary. Also, the human related components were grouped as one single component (human-in-the-loop), which interacts with Machine Actuators, Machine Control, Machine Perception and Environment. A single component represents a more realistic approach facing the complexity added by the human in the system and allows the examination of the user actions and interactions as suggested by [80]. Also, it supports a necessary human-centered and holistic view [81] to better support the complexity of the human behavior and its interaction to the system. It avoids the misconceptions of the too logical designs from some engineering designs and helps to consider and accept human behavior the way it is, not the way engineers would wish it to be [82].

In fact, this is a necessary upgrade considering the original model derived from the classical view of the industrial automation engineering. In those applications, different from the AV systems, the system designer can make some simplifying assumptions: (1) environment can be considered controllable; and (2) the human interactions are simple and have a narrow scope. Furthermore, in many cases, the potential consequences of the human-in-the-loop factor to the whole system can be considered smaller in industrial systems than in AV systems.

As a result, the proposed DDT version (Fig. 5) can be used to map the selected scientific literature. Thus, it can provide a concise perspective on how the field literature covers those main components and which the uncovered areas are. Also, it can provide a good overview on the predominance of the

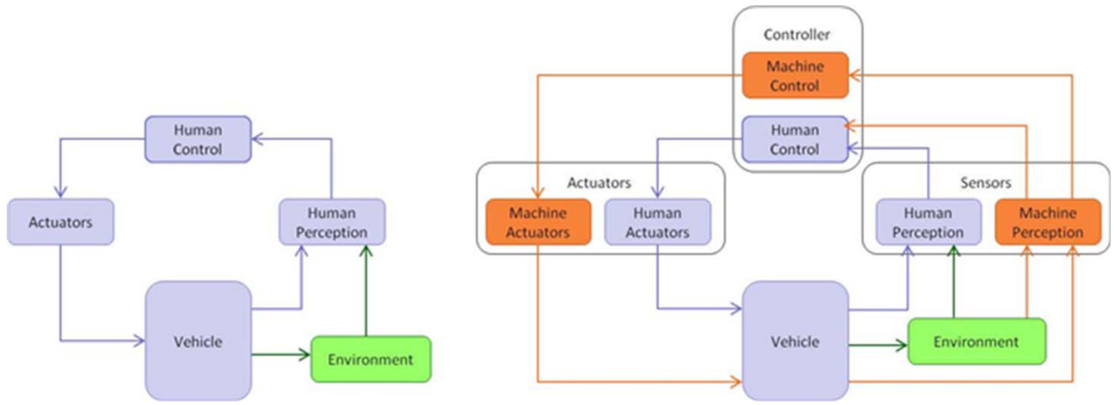

Fig. 4. Models: no automation (left) x semi-automation (right) – source: [79].

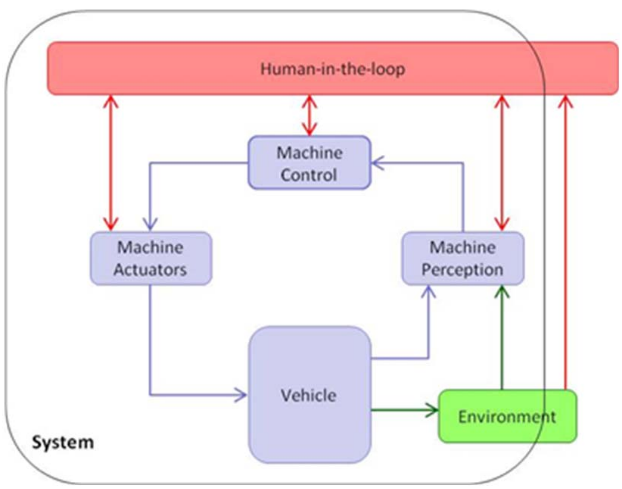

Fig. 5. An adapted version of Dynamic Driving Task (DDT) model.

TABLE V

MODIFIED SEMI-AUTONOMOUS DDT SYSTEM MODEL X CT3 CODES

CT2 - Topic	Component of the Modified DDT System Model (#Hits)	%
CT1 – Impact on Safety: Increase Safety (+)		
Sensors and Perception Fault Prevention	Machine Perception (28)	47
Navigation and Control	Machine Control (11)	19
Human Factor	Human-in-the-loop (5)	8
Conceptual Model and Framework	System (+) (4)	
CT1 – Impact on Safety: Decrease Safety (-)		
Fault Forecasting Ethics and Policies Dependence and Trust	System (-) (11)	25

papers' valence (increase or decrease) on safety. Therefore, the following five subsections details - through the six categories - the SLR findings oriented by these proposed AV system model.

A. CT1 (Impact) and CT2 (Topic)

Considering CT1 (Impact) and CT2 (Topic) codes, Table V shows how they are mapped to the components of the modified semi-autonomous system model, as well as their relationship. Most of the papers are related to machine perception, followed by papers related to a broad system view. Then, the next largest group of papers is related to the machine control component. The remaining papers are related to the human-in-the-loop aspects. An interesting aspect is that only the studies with a broad system aspect were found to have both CT1 codes (increase and decrease system safety). Basically, the studies focused on distinct components solely understand AI can increase the safety risk. Therefore, there is a lack of studies with a critical mindset that explore the potential negative impacts of AI on the individual components. Finally, no papers were found related to the vehicle, machine actuators or environment.

B. CT3 (Techniques)

Considering CT3 code (Techniques), Table VI shows how the wide range of AI techniques are mapped to the components of the modified semi-autonomous system model. The AI techniques are grouped around their scope: system-oriented (19 papers, 32%) and component-oriented (40 papers, 68%). When a paper uses a combination of techniques, for example, ANN and SVM, it results into a unit added to the total number of papers using ANN and a unit added to the total number of papers using SVM. In this context, most of the studies (12 papers, 63%) related to system-wide scope referred to general AI/MI. Most of the studies (11 papers, 20%) related to machine perception used ANNs. In fact, ANN, SVM and HMM (Hidden Markov Model) account for 48% of the studies related to machine perception. Fuzzy logic (3 papers, 18%) is the most widely used technique in the machine control-related papers. Fuzzy Logic, SVM, Optimization Heuristics and Ramer-Douglas-Peucker or Ramer Douglas algorithm account for 53% of the studies related to machine control. Finally, Bayesian Artificial Intelligence techniques are used in most of the studies (29%) related to human-in-the-loop.

Table VII shows the total count of each AI technique occurrence over the sample papers. The sample papers have different heterogeneity in the applied AI approaches. Besides 24% of the papers using generic AI/ML concepts, 49% of the papers applied only one type of AI technique. Therefore,

TABLE VI
TECHNIQUES X DDT SYSTEM MODEL COMPONENT

DDT System Model		Technique	#Hits	%Paper	Accum.%
System-oriented (System)		General AI/ML	12	63%	63%
		Hough Transformation related approaches	2	11%	74%
		Artificial Neural Networks	2	11%	84%
		Optimization Heuristics	1	5%	89%
		Estimation Filters (e.g. Kalman Filter and Particle Filters)	1	5%	95%
		Linear Temporal Logic (LTL)	1	5%	100%
Component-oriented	Machine Perception	Artificial Neural Networks	11	20%	20%
		Support Vector Machine (SVM)	8	15%	35%
		Hidden Markov Based Models (e.g. Continuous Hidden Markov Model-CHMM and Discrete Hidden Markov Model-DHMM)	4	7%	43%
		Estimation Filters (e.g. Kalman Filter and Particle Filters)	3	6%	48%
		Histogram of Oriented Gradient (HOG)	3	6%	54%
		Nearest-Neighbor Based Algorithm (e.g. k-Nearest Neighbors - kNN)	3	6%	59%
		Adaptive Boosting (AdaBoost)	3	6%	65%
		Principal Components Analysis (PCA)	2	4%	69%
		Haar-like feature detector	2	4%	72%
		Fuzzy Logic	2	4%	76%
		Viterbi algorithm	1	2%	78%
		Bayesian Artificial Intelligence (e.g. Bayesian Deep Learning, Naive Bayes Classifier-NBC, etc.)	1	2%	80%
		Regression Based Models	1	2%	81%
		Hough Transformation related approaches	1	2%	83%
		Ramer-Douglas-Peucker or Ramer-Douglas algorithm	1	2%	85%
		Novel Image Recognition Technique	1	2%	87%
		Gaussian Mixture Model (GMM)	1	2%	89%
		General AI/ML	1	2%	91%
		Complex Decision Trees (CDT)	1	2%	93%
		Channel Features	1	2%	94%
		Local Binary Patterns (LBP)	1	2%	96%
		Clustering algorithm k-mean	1	2%	98%
		Conditional Random Fields (CRFs)	1	2%	100%
	Machine Control	Fuzzy Logic	3	18%	18%
		Support Vector Machine (SVM)	2	12%	29%
		Optimization Heuristics	2	12%	41%
		Ramer-Douglas-Peucker or Ramer-Douglas algorithm	2	12%	53%
		Case-based reasoning (CBR)	1	6%	59%
		Nearest-Neighbor Based Algorithm (e.g. k-Nearest Neighbors - kNN)	1	6%	65%
		Basic AI Path Planning algorithms such as A* and D*	1	6%	71%
		Artificial Neural Networks	1	6%	76%
		Regression Based Models	1	6%	82%
		Distributed Random Forest (DRF)	1	6%	88%
		Neuroevolution of Augmenting Topologies (NEAT) - ANN + GA	1	6%	94%
		Satisfiability Modulo Theories (SMT) Solver	1	6%	100%
	Human-in-the-loop	Bayesian Artificial Intelligence (e.g. Bayesian Deep Learning, Naive Bayes Classifier-NBC, etc.)	2	29%	29%
		Regression Based Models	1	14%	43%
		Haar-like feature detector	1	14%	57%
		Canny Edge Detection Algorithm	1	14%	71%
		Hough Transformation related approaches	1	14%	86%
		General AI/ML	1	14%	100%

TABLE VII
HETEROGENEITY OF THE USED AI APPROACHES

Heterogeneity	%	Main Technique	#Hits	%Papers
Generic	24%	General AI/ML	14	100%
Homogenous	49%	Artificial Neural Networks	8	28%
		Fuzzy Logic	4	14%
		Support Vector Machine (SVM)	3	10%
		Regression Based Models	2	7%
		Estimation Filters (e.g. Kalman Filter and Particle Filters)	2	7%
		Bayesian Artificial Intelligence	2	7%
		Optimization Heuristics	2	7%
		Ramer-Douglas-Peucker or Ramer-Douglas algorithm	2	7%
		Hough Transformation	1	3%
		Satisfiability Modulo Theories (SMT) Solver	1	3%
		Adaptive Boosting (AdaBoost)	1	3%
		Linear Temporal Logic (LTL)	1	3%
Hybrid	27%	Artificial Neural Network combined to other techniques	7	44%
		Support Vector Machine (SVM) combined to other techniques	4	25%
		Hidden Markov Based Models (e.g. Continuous Hidden Markov Model-CHMM and Discrete Hidden Markov Model-DHMM) combined to other techniques	2	13%
		Hough Transformation related approaches combined to other techniques	1	6%
		Regression Based Models combined to other techniques	1	6%
		Novel Image Recognition Technique	1	6%

they are homogeneous in terms of the applied AI technique. In those studies, the most widely used techniques were Artificial Neural Networks (8 papers, 28%), Fuzzy Logic (4 papers, 14%) and Support Vector Machine (SVM) (3 papers, 10%). The remaining 27% employed a hybrid approach by combining multiple types of AI techniques. Among those papers, the combination of Artificial Neural Networks to other techniques (7 papers, 44%), Support Vector Machine to other techniques (SVM) (4 papers, 25%) and Hidden Markov-Based Models (e.g. Continuous Hidden Markov Model-CHMM and Discrete Hidden Markov Model-DHMM) to other techniques (2 papers, 13%) were the most frequent hybrid approaches found in the papers selected.

Many different combinations of ANNs with other techniques were found (7 papers). As shown in Table VIII most of those papers are related to Sensors and Perception (3 papers) as well as Navigation and Control (2 papers). Also, papers related to Conceptual Model and Framework and Fault Prevention employed hybrid approach (2 papers). The papers that used a combination of models associated to Hidden Markov Based Models were related to Sensors and Perception. The paper that used Hough Transformation combined to other models is related to Human Factor. The paper that employed a combination of techniques to propose a Novel Image Recognition Technique is related to Sensors and Perception. The paper using Regression-Based Models combined to other techniques is related to AV Navigation and Control. Finally, all the papers employing SVM combined to other techniques were related to the topic Sensors and Perception.

C. CT4 (Problem)

In order to evaluate the CT4 code (Problem) related to the components of the modified semi-autonomous system model, the same grouping strategy applied to Table VI (system-oriented and component-oriented) can be applied here. The *System-level* problems included in 15 papers are: AV Validation [66], Machine-learning-based systems validation to the ultra-dependable levels required for AV [69], Human and Machine Driver Co-existence [62], Coexistence Human Machine Controller [72], Driving Car Tasks Classification [63], Lack of efficient Safety Performance Verification technique when AI/ML is used [67], Crash assignment, especially between automated vehicles and non-automated vehicles [71], Reduce the uncertainty of a driver behavior prediction model [64], Investigate three under-explored themes for AV research: safety, interpretability, and compliance [75], How vehicle autonomy technology can be used to benefit car drivers and also to propose a concept of an autonomous highway vehicle which improves highway driving safety [65], AV decisions in complex dilemmas as a social agent [73], Hybrid (humans and machines) collective decision-making systems [74], Autonomy assurance and trust in Automated Transportation Systems [76], AV Test [68] and, Evaluate the impacts of the number of highly automated vehicles on future traffic safety and traffic flow [70].

Considering the *component-level* problems, 28 papers (47%) are related to dealing with algorithms and techniques to deal with Machine Perception issues, such as: Vehicle Cyber Attack [39], Turn Signal Recognition [33], Securing connected vehicles against Denial of Service (DoS) attack [52], Road Detection [29], Traffic Light Detection [36], Prediction of advanced driver assistance systems (ADAS) remaining useful life (RUL) for the prognosis of ADAS safety critical components [53], Vehicle Detection and Counting [37], predicts the CPU usage patterns of software tasks running on a self-driving car [54], a safety warning and driver-assistance system and an automatic pilot for rural and urban traffic environments [20], reliable and robust obstacles detection

TABLE VIII
HYBRID AI APPROACHES X TOPIC

Main AI Technique	Topic	AI Techniques	Reference
Artificial Neural Network	Conceptual Model and Framework	Hough Transforms, Hough Lines, LocalMaximaFinder, Kalman filters and Convolutional Neural Network (CNN)	[62]
	Fault Prevention	KNN, SVM Regression (SMO), ANN	[53]
	Navigation and Control	CBR, ANN, fuzzy logic, Nearest-Neighbor Retrieval Algorithm, Basic AI Path Planning algorithms such as A* and D*	[39]
		ANN combined to Genetic Algorithm - Neuroevolution of Augmenting Topologies (NEAT)	[48]
	Sensors and Perception	ANNs, AdaBoost, SVM, Hidden Markov Models (HMMs), CRFs	[30]
		Clustering algorithm k- mean, ANN	[31]
		HOG, SVM, PCA, ANN	[28]
Hidden Markov Based Models	Sensors and Perception	GMM, Continuous Hidden Markov Model (CHMM), Discrete Hidden Markov Model (DHMM)	[46]
		HMM, Viterbi algorithm, Adaboost trained Haar-like feature detector	[37]
Hough Transformation	Human Factor	Haar Feature Based Cascade Classifier, Canny edge detection and Hough line transformation	[61]
Novel Image Recognition Technique	Sensors and Perception	Combination of mathematical techniques	[35]
Regression Based Models	Navigation and Control	(DRF) and Linear Regression (LR)	[40]
Support Vector Machine (SVM)	Sensors and Perception	Haar, HOG, LBP, Chanel features, SVM	[38]
		k-Nearest Neighbors (kNN), Naïve Bayes classifier (NBC), SVM	[26]
		Principal component analysis network (PCANet), SVM	[36]
		SVM, HOG	[29]

continues to be largely investigated and still remains an open challenge, especially for difficult scenarios and, in general cases, with loosened constraints and multiple simultaneous use-cases [24], Pedestrian Detection [26], Road environmental recognition and various object detection in real driving conditions [28], Obstacle clustering and tracking [22]. For an autonomous behavior, each truck must be able to follow the vehicle ahead. Due to that, each vehicle must be able to recognize the leading vehicle [21], Speed bump detection [35], providing road safety to connected drivers and connected autonomous vehicles [19], how to ”automate” manual annotation for images to train visual perception for AVs [38], Road Sign Classification in Real-time [32], Road Terrain detection [31], Spatio-temporal situation awareness [34], Pedestrian detection and movement direction recognition [25], Pedestrian Trajectory Prediction [27], Road junction detection [30], Cyber Attack in V2X [55], Learn from Demonstration [46], Early detection of faults or malfunction [56], Road and Obstacle Detection [23], and Enhance Image Understanding [18].

The problems related to Machine Control were found in 11 papers (19%). Those problems include: Pre-Crash problem of Intelligent Control of autonomous vehicles robot [39], Safe-optimal trajectory selection for autonomous vehicle [40], Driverless car 100-km experiment [77], Robot maneuvers too close to an obstacle, which increases the probability of an accident. Preventing this is crucial in dynamic environments, where the obstacles, such as other UAVs, are moving [42], Learning and simulation of the Human-Level decisions involved in driving a racing car [48], Control intersection crossing (all way stop) and optimizing it [43], How to prove the correctness of an algorithm for Vehicle Coordination [44], Path tracking [49], Drivers maneuver classification [45], AVs intersections crossing optimization [47] and Manage low level vehicle actuators (steering throttle and brake) [50].

Finally, the problems related to Human-in-the-loop are present in 5 papers (8%). Those problems include: Selective Attention Mechanism [57], Developing remote controlled car with some automation to deal with traffic light detection, obstacle avoidance system and lane detection system to be driven from anywhere over a secured internet connection [61], Collision avoidance when no action is taken by driver to avoid the collision [58], Human drivers monitoring system to ensure they will be able to take over control within short notice [59] and, Design of driving assistance system [60]. This seems to be an attention-point; this problem category can be considered one serious challenge to semi-autonomous vehicles (SAE Level 1 to Level 4). Therefore, more research is needed into this topic because only 5 papers were found.

D. CT5 (Findings)

Machine Perception has more studies with practical results. Considering the other components, few studies with practical results from real deployments were found. Most of the papers presented preliminary results. In fact, some papers start with a promise and finish with more promises. Considering the total number of papers in this study, only 24% of them were published by journals. Therefore, it is possible to conclude that the field is not mature yet.

Some similar issues were studied in more than one paper about Machine Perception, and distinct techniques were applied to address them (for example, ANN and SVM applied to the topic cyber-attack). Considering some of those techniques have different working mechanisms, that fact can be an important finding for the safety of autonomous cars as regards the need of redundant components.

The papers related to Machine Control also reported positive and promising results, although the level of maturity of the achievements are clearly much lower than the sensors and perception as well as far from what would be expected for an autonomous vehicle considering the potential hazardous situations it may face. In fact, most of the results presented are preliminary.

Only few studies related to human-in-the-loop had practical results from real deployments. However, they seem to be one of the most important topics seeing that there will be more semi-autonomous cars than fully autonomous ones for a while, and they will co-exist. The human factor will thus be an important variable in the system to be considered not only as the impacted side of the safety, but as one of the sources of interactions influencing the safety levels. The topic requires multidisciplinary studies involving fields beyond engineering and computer science, such as neurosciences. This shows the field is not mature yet.

Regarding the system-level, only few of the studies described practical results from real deployments. The papers proposing conceptual models and frameworks bring important contributions, but they are mostly not tested in real set-ups. There is thus a lack of reported results from models and frameworks that could build the foundation of AVs safety.

E. CT6 (Future Works)

A research agenda must consider a serious safety agenda for future studies, at system-level, component-level and AI technique-level. In fact, there are some topics related to safety concerns over AVs, which are critical-path to the development of the field. Some of the suggested topics are related to: the challenges with validating machine-learning-based systems to the ultra-dependable levels required for AVs; wider and deeper studies about human-machine collaboration in the context of AVs; autonomy assurance and trust in AVs; ethical and moral decisions in the context of AVs; among other topics. From them, validating machine-learning-based systems to the ultra-dependable levels required for AVs and autonomy assurance and trust in AVs seem to be the holy grail towards a fully autonomous AV - SAE level 5.

They are also key topics for the Safety Certification of non-deterministic control systems. In those contexts, there are many gaps to be filled by future researches, such as AVs software testing, Fault Injection Testing for AI on AVs, Failure Modes and Effects Analysis (FMEA) for AI on AVs, AI safeguards for AVs, AI safety envelopes for AVs, AI redundancy for AVs (many possible approaches, such as a hybrid connectionist and symbolic architecture using causal inference), explainable AI for AVs, AI fault forecasting.

Finally, studies on V2X communication can help autonomy assurance by providing channels for hardware and software redundancy. Human-machine collaboration in the context of AVs is another key topic with special impact on the semi-autonomous vehicles (SAE levels 1 to 4). Investigations on the best way humans and AVs can interact during normal operations and facing hazardous situations are needed to meet the adequate safety requirements the semi-autonomous vehicles must have. Those studies must consider hybrid collective decision-making systems to enable humans and machines to work together and to agree on common decisions, as well as how to deal with the lack of agreement in some situations.

There is another important discussion arising in the context of human-machine collaboration that must be investigated. On the one hand, there are reports about advanced driver assistant technologies that failed (such as Tesla Autopilot) and the driver was not able to react in time to avoid the accident. They ended-up in life losses and property damages. On the other hand, there are reports about situations in which the advanced driver assistant technologies saved the drivers' life by automatically taking the driver suffering a heart attack to the hospital; fully controlling the car with a drunk driver sleeping; and using a defensive lane change maneuver to avoid being hit by a truck changing its lane. Some players in the industry are pushing the automation evolution steps towards full automation by requiring the human driver to be a backup to the automated driver. Other players in the industry believe the automated driver must be a backup to the human driver. It looks like the second approach can be a smoother and safer path towards SAE level-5 automation.

Immersive environments for training and testing AVs represent another research trend. As the underlying technologies supporting AVs development evolve, higher automation-levels become possible. Considering the potential hazards until the AVs are well trained and fine-tuned, the immersive technologies are becoming an important tool to support the development, training and tests of fully autonomous machines.

Another broad topic requiring further research is related to ethical and moral decisions in AVs. Some studies only mention issues related to moral dilemmas while others provide some simple experiments involving simulated environments and/or human interviews. However, they misinterpret important concepts and bring the discussions around the decisions AVs must make when life losses are involved, besides the moral and ethical perceptions from the human perspective. All of them miss important points such as statistical considerations and the societal result. In other words, the discussions are not deep enough as regards situations such as whether an AV should hit an old man or a child, while a true safe machine control should consider all the probabilities involved and select the one that minimizes the chances of life losses instead of just picking an option. For example, the system must consider small signals, such as which of the potential victims is paying attention to the approaching AV and what would their potential reaction be

TABLE IX
CT3 x CT4

Technique [CT3]	Hits	Papers	Addressed Problem [CT4]	References
General AI/ML	14	24%	AV Validation; Challenge with validating machine-learning based systems to the ultra-dependable levels required for autonomous vehicle; Coexistence Human Machine Controller; Driving Car Tasks Classification; Lack of efficient Safety Performance Verification technique when AI/ML is used; Crash assignment, especially between automated vehicles and non-automated vehicles; Reducing the uncertainty of a driver behavior prediction model; Integration between automatic vehicle and human driver; How the vehicle autonomy technology can be used to benefit car drivers and to improve highway driving safety by a concept of an autonomous highway vehicle; AV decisions in complex dilemmas as a social agent; Hybrid (humans and machines) collective decision making systems (work together and agree on common decisions); Autonomy assurance and trust (CERTIFICATION PROCESS) in Automated Transportation Systems; Evaluating the impacts of the number of highly automated vehicles on future traffic safety and traffic flow; Enhancing Image Understanding.	[18], [59], [63]–[67], [69]–[74], [76]
Artificial Neural Networks	13	22%	Vehicle Cyber Attack; Turn Signal Recognition; Pre- crash issues of Intelligent Control of autonomous ve- hicles robot; Real-time Road Sign Classification; Road Terrain detection; Spatio-temporal situation awareness; Pedestrian detection and movement direction recognition; Pedestrian Trajectory Prediction; Road junction detec- tion; Early faults or malfunction detection; Prediction of advanced driver assistance systems (ADAS) remaining useful life (RUL) for the prognosis of ADAS safety critical components; Road environmental recognition and various objects detection in real driving conditions; Human and Machine Driver Co-existence;	[25], [27], [28], [30]–[34], [39], [53], [56], [62]
Support Vector Machine	10	17%	Road Detection; Robot maneuvers too close to an obstacle; Road environmental recognition and various object detection in real driving conditions; Drivers maneuver classification; Traffic Light Detection; Prediction of adc vanced driver assistance systems (ADAS) remaining useful life (RUL) for the prognosis of ADAS safety critical components Pedestrian Detection; How to ”automate” manual annotation for images to train visual perception for AVs Road junction detection;	[19], [26], [28]–[30], [36], [38], [42], [45], [53]
Bayesian Artificial Intelligence	4	7%	Collision avoidance when no action is taken by driver; Safety, interpretability, and compliance; Pedestrian Detection; Design of driving assistance system;	[26], [58], [60], [75]
Fuzzy Logic	5	8%	PreCrash problem of Intelligent Control of autonomous vehicles robot; Driverless car 100 km experiment Cyber Attack in V2X; Manage low level vehicle actuators (steering throttle and brake); Road and Obstacle Detection;	[23], [39], [50], [55], [77]
Hidden Markov Based Models	4	7%	Vehicle Detection and Counting; Road junction detection; Learn from Demonstration;	[30], [37], [46]
Estimation Filters	4	7%	Human and Machine Driver Co-existence; Securing connected vehicles against Denial of Service (DoS) attack; Reliable and robust obstacles detection;	[24], [52], [62]
Nearest Neighbour-Based Algorithm	4	7%	Pre-crash problem of Intelligent Control of autonomous vehicles robot; Pedestrian Detection; Providing road safety to connected drivers and connected autonomous vehicles;	[19], [26], [39]
Adaptive Boosting	3	5%	Vehicle Detection and Counting; Leading vehicle recogV nition in platooning; Road junction detection;	[21], [30], [37]
Ramer-Douglas Peucker or Ramer-Douglas algorithm	3	5%	Obstacle clustering and tracking; Path tracking;	[22], [49]

TABLE IX
CT3 x CT4

Haar-like feature Detector	3	5%	Developing remote- controlled car with some automation to deal with traffic light detection, obstacle avoidance system and lane detection system to be driven from anywhere over a secured internet connection; Vehicle Detection and Counting; Automating manual annotation for images to train visual perception for AVs;	[37], [38], [61]
Histogram of Oriented Gradient	3	5%	Road Detection; Road environmental recognition and various objects detection in real driving conditions; Automating manual annotation for images to train visual perception for AVs;	[28], [29], [38]
Hough Transformation	3	5%	Road Detection; Road environmental recognition and various object detection in real driving conditions; Automating manual annotation for images to train visual perception for AVs;	[20], [61], [62]
Optimization Heuristics	3	5%	Control intersection crossing (all way stop) and op- timization; Autonomous vehicles intersections crossing optimization; Human and Machine Driver Co-existence;	[43], [47], [62]
Regression- Based Models	3	5%	Selective Attention Mechanism; Safe-optimal trajectory selection for autonomous vehicles; Predicts the CPU usage patterns of software tasks running on a self-driving car;	[40], [54], [57]
Principal Componen-ts Analysis	2	3%	Traffic Light Detection; Road environmental recognition and various object detection in real driving conditions;	[28], [36]
Canny Edge De- tection Algorithm	1	2%	Developing remote- controlled car with some automation to deal with traffic light detection, obstacle avoidance system and lane detection system to be driven from anywhere over a secured internet connection;	[61]
Case-based reasoning	1	2%	Pre-crash problem of Intelligent Control of autonomous vehicles robot;	[39]
Channel Features	1	2%	Automating manual annotation for images to train visual perception for AVs;	[38]
Clustering Algo- rithm k-mean	1	2%	Road Terrain detection;	[31]
Complex Decision Trees	1	2%	Providing road safety to connected drivers and connected autonomous vehicles;	[19]
Conditional Random Fields	1	2%	Road junction detection;	[30]
Distributed Random Forest	1	2%	Safe-optimal trajectory selection for autonomous vehicle;	[40]
Gaussian Mixture Model	1	2%	Learn from Demonstration;	[46]
Linear Temporal Logic	1	2%	AV Test;	[68]
Local Binary Patterns	1	2%	Automating manual annotation for images to train visual perception for AVs;	[38]
Neuroevolution of Augmenting Topologies	1	2%	Learning and simulation of the Human-Level decisions involved in driving a racing car;	[48]
Novel Image Recognition Technique	1	2%	Speed bump detection;	[35]
Path Planning Al- gorithms	1	2%	PreCrash problem of Intelligent Control of autonomous vehicles robot;	[39]
Satisfiability Modulo Theories Solver	1	2%	How to prove the correctness of an algorithm for Vehicle Coordination;	[44]
Viterbi Algorithm	1	2%	Vehicle Detection and Counting;	[37]

and effectiveness of it based on the age and other metrics, as well, considering the multiple scenarios, and the configuration of each, such as speed, region of the car hitting which region of each victim, the potential damages and the severity of the damages considering the estimated weight and overall physical condition, to decide based on the minimization of chances of life losses. This approach will result into higher safety levels for society.

Finally, only 1 paper about autonomous truck was found. Considering some specificities of autonomous truck and its risks, at least a few more studies about the topic could be expected.

V. Conclusion

This paper advanced the literature on the AV by painting a clear picture of the state of the art of the literature in AI on AV safety and a complete view of the positive and negative impacts of AI on AV safety. It was based on initial sample of 4870 retrieved papers, in which 59 studies were selected and mapped into six categories to answer the proposed research questions using a SLR protocol.

Moreover, this study proposed an AV system model. This model extended the DDT model by adding the human-in-the-loop component to the system. This helps a more realistic system representation facing the complexity added by the human in the system. This new model was used to support the literature organization, its analysis and the discussion. Furthermore, a comprehensive mapping of the AI techniques used in the literature was done.

The amplitude and range of the reported future researches in the reviewed papers suggest that there is an empty space for new research into this field. For example, only few studies were found about the three topics positioning AI as a potential source of negative impact on safety - Fault Forecasting, Ethics and Policies, and Dependability and Trust. When combined to the other findings reported by the present study, it confirms the impressions formed during an exploratory research of the literature [1]. It reinforces the perception that the field of AI and AV is not heavily influenced by the safety engineering culture yet. In fact, the studies published about this current topic seem to be more driven by computation-related domains, with no tradition regarding safety culture, than other fields that are much more connected to safety in critical systems [1].

We concluded that additional research is necessary for most of the studies reviewed. They need to be extended to be tested in simulated or real set-ups, new and broader scenarios, with new and more data, and consider experimental designs whereby the results from the proposed approach are compared to benchmarks and alternative techniques. Many AI techniques have achieved impressive results. However, it is still arguable whether the error rates are suitable for real deployments in AVs under the light of a (missing) hazard analysis. Therefore, additional studies with improvements in those techniques are required. Finally, a stronger influence of safety engineering on most of the studies would benefit the research agenda for AI-based AV systems.

Appendix

See Table IX.

References

[1] A. M. Nascimento *et al.*, “Concerns on the differences between ai and system safety mindsets impacting autonomous vehicles safety,” in *Proc. Int. Conf. Comput. Saf., Rel., Secur.*, 2018, pp. 481–486.

[2] E. Scornavacca, S. J. Barnes, and S. L. Huff, “Mobile business research published in 2000-2004: Emergence, current status, and future opportunities,” *Commun. Assoc. Inf. Syst.*, vol. 17, no. 1, p. 28, 2006.

[3] O. Jones and C. Gatrell, “Editorial: The future of writing and reviewing for IJMR,” *Int. J. Manage. Rev.*, vol. 16, no. 3, pp. 249–264, Jul. 2014.

[4] J. Webster and R. T. Watson, “Analyzing the past to prepare for the future: Writing a literature review,” *MIS Quart.*, vol. 26, no. 2, pp. 13–23, 2002.

[5] S. A. Asadollah, D. Sundmark, S. Eldh, H. Hansson, and W. Afzal, “10 Years of research on debugging concurrent and multicore software: A systematic mapping study,” *Softw. Qual. J.*, vol. 25, no. 1, pp. 49–82, 2017.

[6] K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson, “Systematic mapping studies in software engineering,” in *Proc. 12th Int. Conf. Eval. Assessment Softw. Eng.*, 2008, pp. 68–77.

[7] Y. Levy and T. J. Ellis, “A systems approach to conduct an effective literature review in support of information systems research,” *Informing Sci.*, vol. 9, pp. 181–212, Jan. 2006.

[8] J. vom Brocke *et al.*, “Reconstructing the giant: On the importance of rigour in documenting the literature search process,” in *Proc. 17th Eur. Conf. Inf. Syst.*, 2009, pp. 2206–2217.

[9] L. Mathiassen, M. Chiasson, and M. Germonprez, “Style Composition in Action Research Publication,” *MIS Quart.*, vol. 36, no. 2, pp. 347–363, 2012.

[10] B. Kitchenham, “Procedures for performing systematic reviews,” Ph.D. dissertation, Dept. Comput. Sci., Keele Univ., Keele, U.K., 2004.

[11] E. Scornavacca, Jr., and S. J. Barnes, “M-banking services in Japan: A strategic perspective,” *Int. J. Mobile Commun.*, vol. 2, no. 1, pp. 51–66, 2004.

[12] J. Brownlee, “A tour of machine learning algorithms,” *Mach. Learn. Mastery*, to be published.

[13] A. Gani, A. Siddiqa, S. Shamshirband, and F. Hanum, “A survey on indexing techniques for big data: Taxonomy and performance evaluation,” *Knowl. Inf. Syst.*, vol. 46, no. 2, pp. 241–284, 2016.

[14] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, “The WEKA data mining software: An update,” *ACM SIGKDD Explor. Newslett.*, vol. 11, no. 1, pp. 10–18, 2009.

[15] B. Baharudin, L. H. Lee, K. Khan, and A. Khan, “A review of machine learning algorithms for text-documents classification,” *J. Adv. Inf. Technol.*, vol. 1, no. 1, pp. 4–20, 2010.

[16] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 3rd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2009.

[17] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed. San Mateo, CA, USA: Morgan Kaufmann, 2011.

[18] R. L. Simpson, Jr., “Strategic computing computer vision: Taking image understanding to the next plateau,” *Proc. SPIE, Image Understand. Man-Mach. Interface*, vol. 758, Jun. 1987, Art. no. 940067, doi: 10.1117/12.940067.

[19] B. Kihei, J. A. Copeland, and Y. Chang, “Automotive Doppler sensing: The Doppler profile with machine learning in Vehicle-to-Vehicle networks for road safety,” in *Proc. IEEE 18th Int. Workshop Signal Process. Adv. Wireless Commun.*, Jul. 2017, pp. 1–5.

[20] Q. Li, N. Zheng, and H. Cheng, “Springrobot: A prototype autonomous vehicle and its algorithms for lane detection,” *IEEE Trans. Intell. Transp. Syst.*, vol. 5, no. 4, pp. 300–308, Dec. 2004.

[21] W. Woeber, M. Kefer, W. Kubinger, and D. Szuegyi, “Evaluation of daylight and thermal infra-red based detection for platooning vehicles,” in *Proc. 23rd DAAAM Int. Symp. Intell. Manuf. Autom.*, vol. 2, 2012, pp. 719–722.

[22] R. Domínguez, E. Onieva, J. Alonso, J. Villagra, and C. González, “LIDAR based perception solution for autonomous vehicles,” in *Proc. 11th Int. Conf. Intell. Syst. Design Appl.*, Nov. 2011, pp. 790–795.

[23] D. D. Giusto, C. S. Regazzoni, and G. Vernazza, “Multisensorial vision for autonomous vehicle driving,” *Proc. SPIE, Appl. Artif. Intell. VII*, vol. 1095, Mar. 1989, Art. no. 010958, doi: 10.1117/12.969360.

[24] M. Allodi, A. Broggi, D. Giaquinto, M. Patander, and A. Prioletti, “Machine learning in tracking associations with stereo vision and lidar observations for an autonomous vehicle,” in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2016, pp. 648–653.

[25] A. Dominguez-Sanchez, S. Orts-Escolano, and M. Cazorla, “Recognizing pedestrian direction using convolutional neural networks,” in *Advances in Computational Intelligence* (Lecture Notes in Computer Science), vol. 10306, I. Rojas, G. Joya, and A. Catala, Eds. Cham, Switzerland: Springer, 2017.

[26] P. J. Navarro, R. Borraz, C. Fernández, and D. Alonso, “A machine learning approach to pedestrian detection for autonomous vehicles using high-definition 3D range data,” *Sensors (Basel)*, vol. 17, no. 1, p. 18, 2016.

[27] J. Bock, J. Kotte, T. Beemelmanns, and M. Klösges, "Self-learning trajectory prediction with recurrent neural networks at intelligent intersections," in *Proc. 3rd Int. Conf. Vehicle Technol. Intell. Transp. Syst.*, vol. 1, 2017, pp. 346–351.

[28] J. Jeon, S.-H. Hwang, and H. Moon, "Monocular vision-based object recognition for autonomous vehicle driving in a real driving environment," in *Proc. 13th Int. Conf. Ubiquitous Robots Ambient Intell.*, Aug. 2016, pp. 393–399.

[29] K. Kazama, Y. Akagi, P. Raksincharoensak, and H. Mouri, "Fundamental study on road detection method using multi-layered distance data with HOG and SVM," *J. Robot. Mechtron.*, vol. 28, no. 6, pp. 870–877, 2016.

[30] D. Habermann, C. E. O. Vido, F. S. Osório, and F. Ramos, "Road junction detection from 3D point clouds," in *Proc. Int. Joint Conf. Neural Netw.*, Jul. 2016, pp. 4934–4940.

[31] F. Ebadi and M. Norouzi, "Road Terrain detection and Classification algorithm based on the Color Feature extraction," in *Proc. Artif. Intell. Robot.*, Apr. 2017, pp. 139–146.

[32] S. Hamdi, H. Faiedh, C. Souani, and K. Besbes, "Road signs classification by ANN for real-time implementation," in *Proc. Int. Conf. Control, Automat. Diagnosis*, Jan. 2017, pp. 328–332.

[33] K. Yoneda, A. Kuramoto, and N. Suganuma, "Convolutional neural network based vehicle turn signal recognition," in *Proc. Int. Conf. Intell. Inform. Biomed. Sci.*, Nov. 2017, pp. 204–205.

[34] E. Ohn-Bar and M. M. Trivedi, "Are all objects equal? Deep spatio-temporal importance prediction in driving videos," *Pattern Recognit.*, vol. 64, pp. 425–436, Apr. 2017.

[35] W. Devapriya, C. N. K. Babu, and T. Srihari, "Advance driver assistance system (ADAS)—Speed bump detection," in *Proc. IEEE Int. Conf. Comput. Intell. Comput. Res.*, Dec. 2015, pp. 1–6.

[36] Z. Chen and X. Huang, "Accurate and reliable detection of traffic lights using multiclass learning and multiobject tracking," *IEEE Intell. Transp. Syst. Mag.*, vol. 8, no. 4, pp. 28–42, Winter 2016, doi: 10.1109/MITS.2016.2605381.

[37] N. Miller, M. A. Thomas, J. A. Eichel, and A. Mishra, "A hidden Markov model for vehicle detection and counting," in *Proc. 12th Conf. Comput. Robot Vis.*, Jun. 2015, pp. 269–276.

[38] A. M. López *et al.*, "Training my car to see using virtual worlds," *Image Vis. Comput.*, vol. 68, pp. 102–118, Dec. 2017.

[39] B. M. Elbagoury, A. M. Salem, and L. Vladareanu, "Intelligent adaptive precrash control for autonmous vehicle agents (CBR Engine & hybrid A∗ path planner)," in *Proc. Int. Conf. Adv. Mechatronic Syst.*, Nov./Dec. 2016, pp. 429–436.

[40] H. A. Najada and I. Mahgoub, "Autonomous vehicles safe-optimal trajectory selection based on big data analysis and predefined user preferences," in *Proc. IEEE 7th Annu. Ubiquitous Comput., Electron. Mobile Commun. Conf.*, Oct. 2016, pp. 1–6.

[41] J. Godoy, J. Villagrá, T. de Pedro, and R. Galán, "Virtual vehicle approach for longitudinal control in Urban environments," in *Computer Aided Systems Theory—EUROCAST 2013* (Lecture Notes in Computer Science), vol. 8112, R. Moreno-Díaz, F. Pichler, and A. Quesada-Arencibia, Eds. Berlin, Germany: Springer, 2013.

[42] O. Linda, T. Vollmer, and M. Manic, "SVM-inspired dynamic safe navigation using convex hull construction," in *Proc. 4th IEEE Conf. Ind. Electron. Appl.*, May 2009, pp. 1001–1006.

[43] I. Zohdy and H. Rakha, "Optimizing driverless vehicles at intersections," in *Proc. 10th ITS World Congr.*, 2012, pp. 1–12.

[44] M. Asplund, A. Manzoor, M. Bouroche, S. Clarke, and V. Cahill, "A formal approach to autonomous vehicle coordination," in *Formal Methods* (Lecture Notes in Computer Science), vol. 7436, D. Giannakopoulou and D. Méry, Eds. Berlin, Germany: Springer, 2012.

[45] E. J. L. Pulgarin, G. Herrmann, and U. Leonards, "Drivers' Manoeuvre classification for safe HRI," in *Towards Autonomous Robotic Systems* (Lecture Notes in Computer Science), vol. 10454, Y. Gao, S. Fallah, Y. Jin, and C. Lekakou, Eds. Cham, Switzerland: Springer, 2017, pp. 475–483.

[46] F. Ge, W. Moore, and M. Antolovich, "Learning from demonstration using GMM, CHMM and DHMM: A comparison," in *Advances in Artificial Intelligence* (Lecture Notes in Computer Science), vol. 9457, B. Pfahringer and J. Renz, Eds. Cham, Switzerland: Springer, 2015.

[47] I. H. Zohdy, R. K. Kamalanathsharma, and H. Rakha, "Intersection management for autonomous vehicles using iCACC," in *Proc. 15th Int. IEEE Conf. Intell. Transp. Syst.*, Sep. 2012, pp. 1109–1114.

[48] F. Gallego, F. Llorens, M. Pujol, and R. Rizo, "Learning human-level AI abilities to drive racing cars," in *Proc. Conf. Artif. Intell. Res. Develop.*, 2005, pp. 33–40.

[49] C. G. Serna, A. Lombard, Y. Ruichek, and A. Abbas-Turki, "GPS-based curve estimation for an adaptive pure pursuit algorithm," in *Advances in Computational Intelligence* (Lecture Notes in Computer Science), vol. 10061, S. Sidorov and O. Herrera-Alcántara, Eds. Cham, Switzerland: Springer, 2017.

[50] J. E. Naranjo, F. Jiménez, O. Gómez, and J. G. Zato, "Low level control layer definition for autonomous vehicles based on fuzzy logic," *Intell. Autom. Soft Comput.*, vol. 18, no. 4, pp. 333–348, 2012.

[51] A. Avizienis, J.-C. Laprie, B. Randell, and C. Landwehr, "Basic concepts and taxonomy of dependable and secure computing," *IEEE Trans. Depend. Sec. Comput.*, vol. 1, no. 1, pp. 11–33, Jan./Mar. 2004.

[52] P. Sharma, H. Liu, H. Wang, and S. Zhang, "Securing wireless communications of connected vehicles with artificial intelligence," in *Proc. IEEE Int. Symp. Technol. Homeland Secur.*, Apr. 2017, pp. 1–7.

[53] M. A. Taie, E. M. Moawad, M. Diab, and M. ElHelw, "Remote diagnosis, maintenance and prognosis for advanced driver assistance systems using machine learning algorithms," *SAE Int. J. Passenger Cars-Electron. Elect. Syst.*, vol. 9, pp. 114–122, Apr. 2016.

[54] Y.-W. Seo, J. Kim, and R. Rajkumar, "Predicting dynamic computational workload of a self-driving car," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Oct. 2014, pp. 3030–3035.

[55] S. Noei, A. Sargolzaei, A. Abbaspour, and K. Yen, "A decision support system for improving resiliency of cooperative adaptive cruise control systems," *Procedia Comput. Sci.*, vol. 95, pp. 489–496, Jan. 2016.

[56] N. R. M. E. Gaura, "Microsensors, arrays and automatic diagnosis of sensor faults," in *Proc. Nanotechnol. Conf. Trade Show*, 2003, vol. 1, no. 1, pp. 276–279.

[57] Q. Guan, H. Bao, and Z. Xuan, "The research of prediction model on intelligent vehicle based on driver's perception," *Cluster Comput.*, vol. 20, no. 4, pp. 2967–2979, 2017.

[58] A. M. Khan, "Design of real-time transition from driving assistance to automation function: Bayesian artificial intelligence approach," in *Advanced Microsystems for Automotive Applications* (Lecture Notes in Mobility), J. Fischer-Wolfarth and G. Meyer, Eds. Cham, Switzerland: Springer, 2014.

[59] H. Rahman, S. Begum, and M. U. Ahmed, "Driver monitoring in the context of autonomous vehicle," in *Proc. 13th Scand. Conf. Artif. Intell.*, 2015, pp. 1–10.

[60] A. Khan, "Cognitive vehicle design guided by human factors and supported by Bayesian artificial intelligence," in *Advances in Human Aspects of Transportation* (Advances in Intelligent Systems and Computing), vol. 597, N. Stanton, Ed. Cham, Switzerland: Springer, 2018.

[61] M. R. T. Hossai, M. A. Shahjalal, and N. F. Nuri, "Design of an IoT based autonomous vehicle with the aid of computer vision," in *2017 Int. Conf. Electr., Comput. Communication Eng. (ECCE)*, vol. 2017, pp. 752–756.

[62] P. K. Gurudatt and V. Umesh, "Novel architecture for cloud based next gen vehicle platform—Transition from today to 2025," in *Proc. IEEE Int. Conf. Consum. Electron.-Asia*, Oct. 2017, pp. 151–155.

[63] S. Ionita, "Autonomous vehicles: From paradigms to technology," in *Proc. IOP Conf. Ser., Mater. Sci. Eng.*, 2017, vol. 252, no. 1, Art. no. 012098.

[64] A. Rakotonirainy, O. Orfila, and D. Gruyer, "Reducing driver's behavioural uncertainties using an interdisciplinary approach: Convergence of quantified self, automated vehicles, Internet of Things and artificial intelligence," *IFAC-PapersOnLine*, vol. 49, no. 32, pp. 78–82, 2016.

[65] A. Okuno, A. Kutami, and K. Fujita, "Towards autonomous cruising on highways," *SAE Trans.*, vol. 99, no. 6, pp. 1343–1352, 1990.

[66] P. Koopman, "Challenges in autonomous vehicle validation: Keynote presentation abstract," in *Proc. 1st Int. Workshop Safe Control Connected Auton. Vehicles*, Apr. 2017, p. 3.

[67] D. Meltz and H. Guterman, "RobIL—Israeli program for research and development of autonomous UGV: Performance evaluation methodology," in *Proc. IEEE Int. Conf. Sci. Elect. Eng.*, Nov. 2016, pp. 1–5.

[68] C. Wolschke, T. Kuhn, D. Rombach, and P. Liggesmeyer, "Observation based creation of minimal test suites for autonomous vehicles," in *Proc. IEEE Int. Symp. Softw. Rel. Eng. Workshops*, Oct. 2017, pp. 294–301.

[69] P. Koopman and M. Wagner, "Autonomous vehicle safety: An interdisciplinary challenge," *IEEE Intell. Transp. Syst. Mag.*, vol. 9, no. 1, pp. 90–96, Jan. 2017.

[70] M. Bahram, Z. Ghandeharioun, P. Zahn, M. Baur, W. Huber, and F. Busch, "Microscopic traffic simulation based evaluation of highly automated driving on highways," in *Proc. 17th Int. IEEE Conf. Intell. Transp. Syst.*, Oct. 2014, pp. 1752–1757.

[71] W. Kumfer and R. Burgess, "Investigation into the role of rational ethics in crashes of automated vehicles," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2489, pp. 130–136, 2015.
[72] G. Lugano, "Virtual assistants and self-driving cars," in *Proc. 15th Int. Conf. ITS Telecommun.*, May 2017, pp. 1–5.
[73] E. Dogan, R. Chatila, S. Chauvier, K. Evans, P. Hadjixenophontos, and J. Perrin, "Ethics in the design of automated vehicles: The AVEthics project," in *Proc. EDIA@ECAI*, 2016, pp. 10–13.
[74] J. Greene, F. Rossi, J. Tasioulas, K. B. Venable, and B. Williams, "Embedding ethical principles in collective decision support systems," in *Proc. 13th AAAI Conf. Artif. Intell.*, 2016, pp. 4147–4151.
[75] R. McAllister *et al.*, "Concrete problems for autonomous vehicle safety: Advantages of Bayesian deep learning," in *Proc. 26th Int. Joint Conf. Artif. Intell.*, vol. 17, 2017, pp. 4745–4753.
[76] R. A. Lacher Grabowski and S. Cook, "Autonomy, trust, and transportation," in *Proc. AAAI Spring Symp. Ser.*, 2014, pp. 42–49.
[77] J. Godoy, J. Pérez, E. Onieva, J. Villagrá, V. Milanés, and R. Haber, "A driverless vehicle demonstration on motorways and in urban environments," *Transport*, vol. 30, no. 3, pp. 253–263, 2015.
[78] A. M. Nascimento *et al.*, "A systematic literature review about the impact of artificial intelligence on autonomous vehicle safety," Apr. 2019, *arXiv:1904.02697*. [Online]. Available: https://arxiv.org/abs/1904.02697
[79] A. Christensen *et al.*, "Key considerations in the development of driving automation systems,"in *Proc. 24th Enhanced Saf. Vehicles Conf.*, 2015, pp. 1–14.
[80] N. R. Storey, *Safety-Critical Computer Systems*. Boston, MA, USA: Addison-Wesley, 1996.
[81] G. Salvendy, *Handbook of Human Factors and Ergonomics*. Hoboken, NJ, USA: Wiley, 2012.
[82] D. Norman, *The Design of Everyday Things: Revised and Expanded*, ed. New York, NY, USA: Basic Books, 2013.

Alexandre Moreira Nascimento received the B.Eng. degree in mechatronics engineering from the School of Engineering, University of São Paulo (Poli-USP), in 1998, the M.Sc. degree in management with major on information system from the School of Management, University of São Paulo (FEA-USP), in 2005, and the M.Sc. degree in management with major on system dynamics from the Massachusetts Institute of Technology (MIT) in 2013. He was with Massachusetts Institute of Technology (MIT) Media Laboratory for an academic term. He was with Samsung Research and Development, Silicon Valley, as an International Product Engineering Director for Samsung Internet of Things (IoT) Platform. He also holds specializations from MIT, Berkeley University, and Harvard University, in topics, such as cognitive robotics, deep learning, image recognition, and text mining. He is currently a Researcher in artificial intelligence and safety on autonomous devices with the Safety Analysis Group, Poli-USP. He is also a Visiting Researcher with Stanford University.

Lucio Flavio Vismari received the B.Eng. and M.Eng. degrees from the School of Engineering, University of São Paulo (Poli-USP), in 2001 and 2007, respectively, all in electrical engineering. Since 2002, he has been with the Safety Analysis Group (GAS), Poli-USP as a Senior Researcher in system safety. His current research interests include the safety assurance challenges in complex engineered systems, mainly the application of cyber-physical systems and new technological trends (as AI, 5G communication, and XR) on high-risk critical domains. He has published more than 40 scientific works in periodicals, conferences, and books. Besides, he has more than 15 years of experience as an independent safety assessor, involved in dozens of large-scale public and private projects in railway, subway, ATM, and defense domains.

Caroline Bianca Santos Tancredi Molina received the B.Sc. degree in science and technology from the Federal University of São Paulo (UNIFESP) in 2013 and the M.Sc. degree in computer engineering from the School of Engineering, University of São Paulo (Poli-USP), in 2007. She is currently pursuing the degree in computer engineering with UNIFESP and the D.Sc. degree with the School of Engineering, Poli-USP. Her D.Sc. research focus is on V2X communications, supported by Ericsson Research.

Paulo Sergio Cugnasca received the M.Sc. and Ph.D. degrees from the University of São Paulo—Poli-USP, Brazil, in 1993 and 1999, respectively. He was an Electronic Engineer with the School of Engineering, Poli-USP, in 1987, where he is currently an Assistant Professor with the Computer Engineering and Digital Systems Department, and is also a member of the Safety Analysis Group. His research interest includes reliable and safe computational systems for critical applications.

João Batista Camargo, Jr. (M'14) received the bachelor's degree in electronic engineering and the M.Sc. and Ph.D. degrees from the School of Engineering, University of São Paulo (Poli-USP), São Paulo, Brazil, in 1981, 1989, and 1996, respectively. He is currently an Associate Professor with the Department of Computer and Digital Systems Engineering (PCS), Poli-USP, where he is also the Coordinator of the Safety Analysis Group (GAS). He has 30 articles in scientific journals, five organized books, six published book chapters, and 115 complete works published in proceedings of conferences. He is a Reviewer in different scientific journals, such as the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, *Journal of Intelligent and Robotic Systems*, *Risk Analysis*, the *Journal of Advanced Transportation, Reliability Engineering and System Safety*, *Transportation Research - Part C*, and the IEEE SOFTWARE.

Jorge Rady de Almeida, Jr., received the M.Sc. and Ph.D. degrees from Poli-USP, Brazil, in 1989 and 1995, respectively. He was an Electronic Engineer with the School of Engineering, Poli-USP, in 1981. He is currently an Associate Professor with the Computer and Digital Systems Engineering Department, University of São Paulo (Poli-USP), Brazil. He is also a member of the Safety Analysis Group at Poli-USP. His research interest includes reliable and safe computational systems for critical application.

Rafia Inam received the Ph.D. degree from Mälardalen University, Sweden, in 2014. She is currently a Senior Project Manager with the Ericsson Research, Sweden, in artificial intelligence. She joined Ericsson Research in 2014 and involved in 5G for industries. She was also a Single Point of Contact with Scania. Her research interests include 5G network slices and management, using AI for automation, use cases applied to 5G, service modeling for intelligent transport systems, automation and safety for CPS, and reusability of real-time software and ITS. She has coauthored more than 38 scientific publications and more than 38 patent families.

Elena Fersman received the M.Sc. degree in economics and management from St. Petersburg Polytechnic University and the Ph.D. degree in computer science from Uppsala University. She is currently a Research Director in artificial intelligence with Ericsson Research. She is responsible for a team of 100+ researchers located in Sweden, U.S., India, Hungary, and Brazil. She is also a Docent and an Adjunct Professor in cyber-physical systems specialized in automation with the Royal Institute of Technology, Stockholm.

She did the Post-Doctoral Research with the University Paris-Saclay. At Ericsson, she had various positions ranging from product management to research leadership. Since 2019, she has been a member of the Board of Directors of RISE Research Institutes of Sweden. She has coauthored over 50 patent families.

Maria Valeria Marquezini received the master's and D.Sc. degrees from the State University of Campinas- UNICAMP, Brazil, in 1990 and 1995, respectively, all in physics. From 1995 to 1997, she was a Post-Doctoral Researcher with the Lawrence Berkeley National Laboratory (LBNL), Materials Science Division, conducting research activities in the field of electronic processes in semiconductors. From 1997 to 1999, she was an Associate Researcher with the Physics Institute, UNICAMP. Since 2000, she has been with the Research Department of Ericsson, Brazil, and has been engaged in many different research projects in the telecommunications domain, ranging from fixed to wireless technologies, and devices to networks technologies. She is currently responsible for coordinating research projects developed in partnership with universities and research institutes in Brazil. She is also a Project Manager from Ericsson side in the H2020 5G Range project, where Ericsson is leading one of the workpackages.

Alberto Yukinobu Hata received the M.Sc. degree in computer science and the Ph.D. degree from the University of São Paulo (ICMC/USP), in 2010 and 2016, respectively. He is currently a Researcher in artificial intelligence with Ericsson Telecommunication S.A., Brasil. He has been working actively and published over 27 articles in the field of robotics, image processing, machine learning, and intelligent transportation systems. His current research interest includes the human–robot collaboration for smart manufactories.