Autonomous Vehicles that Cooperate and Understand

Intelligent Algorithms under the Hood

Jiří Wiedermann

Jan van Leeuwen

Technical Report UU-PCS-2021-01
January 2021

Center for Philosophy of Computer Science
Department of Information and Computing Sciences
Utrecht University, Utrecht, The Netherlands
www.cs.uu.nl

Series: UU-PCS

Department of Information and Computing Sciences
Utrecht University
Princetonplein 5
3584 CC Utrecht
The Netherlands

Autonomous Vehicles that Cooperate and Understand: Intelligent Algorithms under the Hood *

Jiří Wiedermann[1] and Jan van Leeuwen[2]

[1] Institute of Computer Science of Czech Academy of Sciences and Karel Čapek Center for Values in Science and Technology, Prague, Czech Republic
jiri.wiedermann@cs.cas.cz
[2] Dept. of Information and Computing Sciences, Utrecht University, the Netherlands
J.vanLeeuwen1@uu.nl

Abstract. We present a new paradigm for the research and development of autonomous vehicles from a philosophical viewpoint. The approach takes the broader epistemic context into account in which the vehicles operate and utilizes cognitive mechanisms inspired by higher-level mental processes. The goal is to design (connected) autonomous vehicles that provably understand all traffic situations that they can perceive and that cooperate with an informed smart road infrastructure in resolving them. The features are seen as a cognitive attainment – the vehicle's ability to handle traffic situations in a way aligned with its mission, based on information about the surrounding traffic and about the past actions of the vehicle. A key ingredient of the approach is to view an autonomous vehicle as a cognitive cyber-physical human system endowed with so-called 'minimal machine consciousness', a prerequisite of machine understanding. Its on-board sensors and the external smart road infrastructure must provide a vehicle with the information that is sufficient to provably elicit its understanding of the evolving traffic situations and fulfil its mission, in cooperation with other vehicles and the smart road infrastructure. We show how the approach leads to a driving algorithm that is arguably safe and reliable for guiding a connected autonomous vehicle to its destination. We discuss the potential of the new paradigm to overcome the difficult issues in autonomous driving.

"Self-driving cars are the natural extension of active safety and obviously something we should do."

Elon Musk, 2013

Keywords: autonomous vehicles, driving algorithm, machine consciousness, machine understanding, philosophy of computing, safety, smart roadside infrastructure.

1 Introduction

Industry heavily invests in the development of autonomous vehicles that have 'level 5 autonomy', i.e. that can drive and reach their destination with minimal or no intervention from a human driver. Achieving it will be a feat of cooperative AI, algorithm design and automotive engineering, with advanced sensor and vision technology feeding intelligent algorithms in and around the vehicle.

Since the first fatal accidents with driver-less cars, however, it has become clear that achieving fully self-driving cars is harder than originally anticipated. The prevailing approach to the design of autonomous vehicles and their driver assistance software,

* Version dated: January 28, 2021. The research of the first author was partially supported by ICS AS CR fund RVO 67985807, programme Strategy AV21 "Hopes and Risks of the Digital Age", and the Karel Čapek Center for Values in Science and Technology.

based on on-board sensors and LIDAR technology and computing by deep neural networks, seems to have reached its technological limits. New ideas have been sought to overcome the apparent 'AI roadblock', as it is called in [4]. The question has arisen in the design of many other robotic systems as well (cf. [22, 23]).

One of the alternative approaches that have been proposed is the idea of increasing the *cognitive abilities* of autonomous vehicles, by endowing them with 'machine versions' of selected higher-level human abilities (cf. [5, 7, 22]). For example, it has been suggested that full autonomy requires self-driving cars and similar systems to be 'self-aware' (cf. [6]). In this paper we develop a complete paradigm for the development of fully autonomous vehicles based on this approach, by applying recent insights into the design of robotic systems from the philosophy of computing [22, 23].

Fully autonomous vehicles will not be completely independent (or, autonomous) from their users. Indeed, complete autonomy of driver-less vehicles should, in general, not be the main goal; rather, the goal should be the development of autonomous vehicles that operate and cooperate purposefully with each other, with their environment, and with humans, for the sake of maximal safety and reliability of their operation (cf. [10]). Thus, understandably, research is heavily focused on the development of *connected autonomous vehicles* (CAVs) [7].

From now on we assume all autonomous vehicles to be connected, i.e. communicating with the other vehicles that it may encounter on its course and to the surrounding infrastructure consisting of traffic signs and roadside sensors.

Cognitive machines? When looking for the causes of the difficulties with today's autonomous vehicles through the prism of their cognitive abilities, one inevitably comes to one conclusion: it is their limited ability to 'understand' and 'manage' general traffic situations that does not allow them to behave in a manner appropriate for such situations. This suggests that future autonomous vehicles should be perfected in their ability to be conscious of their state and surroundings, understand the traffic situations they encounter, operate in them, and influence them to a desired effect in cooperation with other vehicles, all as part of reaching their goal efficiently and safely.

Clearly, these ideas must be transferred to the automotive context in an appropriate way. For example, 'understanding' is a philosophical notion (cf. [2], but when it is applied to autonomous vehicles, we rather speak of 'machine understanding'. We then see it as an effective relation between the *subject* that understands – the autonomous car in its given cognitive state, and the *object* of its understanding – the traffic situation it is in. This relation must provably imply abilities and dispositions of the vehicle with respect to the object of understanding (the traffic situation) that are sufficient to handle it in a way that is aligned with the purpose of the vehicle (cf. [25]). Many other higher-level cognitive abilities can be transfered to this context too, and are heavily studied in AI from a computational viewpoint as well. For an overview of the recent status quo in this field, we refer to [5, 7, 14, 15, 19, 22].

Can one approach the design of connected autonomous vehicles from a broader epistemic and cognitive perspective and use it to give a new impetus to the field? This leads to the main challenge we address in this paper:

> *Can one build the concept of safe and reliable (connected) autonomous vehicles around (machine versions of) the core notions from cognition like 'consciousness', 'cooperation', and 'understanding'?*

In other words: can connected autonomous vehicles be viewed as cognitive machines, and can a design philosophy for them be based on it? This would extend our approach to general robotic systems in [22], and generalize the ideas put forth in recent studies like [5, 6].

Paradigm In this paper we meet the challenge, by taking a fresh look at the design of autonomous vehicles and the systems that guide them. We will present a new *paradigm* for their development, based on far-reaching machine analogies to the principles and concepts of cognitive systems. The resulting framework makes it possible to link to the current trends in the fields of software engineering, artificial intelligence and robotics. The ultimate aim is to enable autonomous vehicles to 'understand' and 'act safely in' all situations in the traffic they encounter and 'cooperate' with each other and the traffic infrastructure towards achieving their goal.

The paradigm we develop is described by means of a *baseline* and *three postulates* and can be characterized as follows:

- the baseline asserts that connected autonomous vehicles must be seen as *cognitive* cyber-physical human systems, and
- the three postulates aim at the design targets that must be minimally realized to enable the vehicles to navigate and advance in a fully self-controlled manner and to reach their set goal. The postulates express that autonomous systems should, respectively, be: *(1)* 'minimally machine conscious' of themselves and the environment, *(2)* fully integrated with a 'smart' roadside infrastructure, and *(3)* endowed with 'machine understanding' for dealing with traffic situations.

Whereas the baseline implies that autonomous vehicles must have the characteristics and architecture of 'cognitive machines', the three postulates focus on their cognitive functionalities. We will argue that the postulates are necessary prerequisites for autonomous vehicles to implement the desired controls towards safety and reliability. We also show that, based on the paradigm, an 'intelligent' *driving algorithm* for autonomous vehicles can be designed that enables the vehicles to reach their goal safely and effectively.

Contributions We advance several contributions towards the philosophy and design of autonomous vehicles:

(i) we show that the design of autonomous vehicles can be seen as an instance of the design of self-controlled and safe cyber-physical human systems. The postulate expressing that the vehicles shall be 'minimally machine conscious' is a direct consequence of, what has been proclaimed to be, the 'manifesto' for the design of these systems [23].

(ii) we argue that 'smart roadside infrastructures' should be designed that not only support, but also 'extend' the cognitive abilities of autonomous vehicles and facilitate their cooperation. The smart roadside infrastructures for guiding autonomous vehicles should thus be developed as integral parts of their control systems.

(iii) we provide a philosophical basis for endowing autonomous vehicles with other higher-level cognitive abilities, notably with 'machine understanding' and its application to the understanding of traffic situations. The corresponding postulate thus has a firm grounding in the general theories of understanding by machines [19]. It will turn out that for autonomous vehicles, minimal machine consciousness is a necessary condition for machine understanding.

The driving algorithm of a minimally conscious autonomous car that cooperates and understands, based on these ideas and proposed in this paper, makes use of the *epistemic approach* to computation [20]. In this approach, computation is viewed as knowledge generation in the framework of a suitable epistemic theory. The application of all these findings hopefully helps to give the field a further impulse towards fulfilling its mission: driver-less riding under all circumstances.

Overview The paper is organized as follows. In Section 2 we present the baseline of our paradigm, viewing connected autonomous vehicles as cognitive cyber-physical human systems. We outline the typical architecture and feedback loops in these systems. Next, we present the three postulates: the one on minimal machine consciousness in Section 3, the one on communication and cooperation using a smart roadside infrastructure in Section 4, and the one on machine understanding in Section 5.

A skeletal driving algorithm for a minimally machine conscious (connected) autonomous vehicle that operates according to our paradigm, is sketched in Section 6. In Section 7 we discuss the effect of the postulates on future design methodologies for CAVs. In Section 8 we present some conclusions.

The framework we develop seems suitable for the further study of formalizations and algorithms pertaining to the design of autonomous vehicles.

Abbreviations AV = autonomous vehicle, CAV = connected autonomous vehicle, CPHS = cyber-physical human system, CCPHS = cognitive cyber-physical human system, MMC = minimally machine conscious, SACA = Sense-Analyse-Compute-Act, SRI = smart roadside infrastructure.

2 Autonomous Vehicles are Cognitive Cyber-Physical Human Systems

We consider autonomous vehicles as they should ideally operate in the physical world, with minimal intervention from a human driver. We assume that the AVs are connected by wireless communication, to each other and to the roadside infrastructure of an 'intelligent traffic system' that helps them to navigate to their destination and drive safely, without collisions. (The vehicles may also be connected to a remote control center, to keep track of them and possibly modify their mission.)

Our aim is to develop a paradigm for the design of CAVs that views them as *cognitive machines*, i.e., as vehicles that can be equipped with machine versions of the cognitive abilities normally ascribed to human drivers to achieve 'full autonomy'. We will focus here especially on the prerequisites for understanding and managing traffic situations. In this section we first define the baseline of the paradigm, describing CAVs as cyber-physical (human) systems. In later sections we will present the three postulates of the paradigm that focus on the cognitive elements.

2.1 Baseline

From a philosophical point of view, CAVs are just instances of a broader class of machines known as cyber-physical systems. A *cyber-physical system* [18] is a robotic system that operates in the real world and that is controlled or monitored by computer-based algorithms and, possibly, by human interventions.

A *cognitive cyber-physical human system* [23] is a CPHS that is capable of 'perception' and that has various cognitive abilities, such as the categorization of perceived

objects and machine versions of concepts like consciousness, understanding and possibly some other. We keep humans 'in the loop', as this is the framework that applies here. The baseline of our paradigm *revises* the concept of CAVs as follows:

Baseline (CAVs are Cognitive Cyber-Physical Human Systems)
A CAV is a cognitive cyber-physical human system *whose purpose is to transport people or goods with little or no human input, safely, reliably, and efficiently between two or more points on a given road map.*

A consequence of the baseline is that CAVs may be designed with the characteristics of cognitive CPHSs in mind. These systems have both the architectural and operational qualities to allow for machine versions of cognitive functions. We discuss the architectural implications and the typical feedback loops of cognitive CPHSs in Section 2.2. Applying this to CAVs, we prepare for the philosophical *postulates* for their design towards full autonomy in Section 2.3.

2.2 Architecture

Cognitive CPHSs are CPHSs which are able to utilize their hard- and software capabilities to act and react *awarely* and with some degree of *(machine) intelligence* with regard to both their own operation and their environment. This is commonly facilitated by augmenting their architecture with networks of sensory gadgets and operator panels, to monitor the system's components and their interfaces and to control the interaction of the CPHS with any actors around it, locally and globally.

In a typical cognitive CPHS, the signals and reports of the sensory networks are continually combined and processed in one or more processing units, to determine and refresh the 'cognitive state' of the system and to compute or steer its sequence of actions. The sensory abilities are a prerequisite for creating the cognitive abilities, and thus of a high-level *supervisory* system that utilizes them, for monitoring and controlling any CPHS in operation and thus, by the baseline, of any CAV.

Feedback CPSs are controlled with the help of *feedback loops* that guarantee a permanent exchange of information and commands between components and control units as the system operates. In cognitive CPHSs, there are special feedback loops based on 'cognitive data' (see below). These loops provide the driving input for the cognitive supervisory system of a cognitive CPHS.

Feedback in a (cognitive) CPHS is typically organized as follows. First, the sensory units continually supply the control unit(s) with *representations* of the perceived 'sensations' for which they are designed. They also provide *feedback signals*, to reflect the 'accuracy' of every reported sensation. The accuracy is normally 'graded' according to some scale depending on the nature of the reported sensations. For example, it can refer to magnitude, intensity, frequency, shapes, and so on.

Similarly, the sensations and feedbacks from the motor units of a cognitive CPHS are supplied in the form of *reports* which must state whether, or to what extent, the intended operations could be realized by a unit. Together, the 'feedback accuracy' and the 'reports' constitute the *quality* of the respective feedback.

Clearly the information flow in a CCPHS is not only directed from the sensory and motor units to the control unit(s) but also vice versa, from the control unit(s) to the

sensory and motor units. In the latter case, the control unit(s) send 'activation signals' to the sensory units and instructions to the motor units continually, as determined by the program of the CPHS.

Sensory networks Seeing CAVs as cognitive CPHSs requires that they can facilitate a cognitive supervisory system as described, with its feedback loops, in order that desired cognitive mechanisms can be supported. The extent and quality of the feedback loops in a cognitive CPHS, and thus in a CAV, ultimately determine how close the systems can come to full autonomy.

One recognizes that, indeed, present-day CAVs deploy a host of sophisticated sensory devices to achieve this goal, from devices like cameras, radars, lidars, various kinds of receivers (GPS, wifi, bluetooth, voice) and touch screens, to sensory units embedded in the vital components of the vehicle like the engine, the wheels, and any of the controlling processors.

It is a consequence of the baseline that the sensory networks in CAVs must be designed as in cognitive CPHS, but geared to the needs and requirements of (machine versions of) the cognitive functions one wants to have for operating the systems under 'full autonomy'. The graded feedback from the sensory and motor units to the control unit is crucial for the operation of cognitive CPHSs, and thus of CAVs.

2.3 Cognitive layer

With the feedback loops in mind, cognitive CPHSs will continually work to maintain their so-called *cognitive state.* This state describes the set of values of all important variable aspects of the system, including the signals received from all sensory and motor units, the signals sent to all its sensory and motor units, the relevant processor states and memories in the system, and the state of the control inputs and outputs to human operators and users. The cognitive state of the system is broadcast to all its modules, to keep the whole system informed.

The programs of cognitive CPHSs, and thus of CAVs, will reflect this cyclic maintenance of their cognitive state. A typical *operational cycle* of the programs will consist of the following four phases, iterated in sequel: *Sense-Analyze-Compute-Act* (SACA), with a meaning similar to e.g. Boyd's *OODA loop* or the *MAPE-K loop* in self-adaptive autonomic systems. The SACA-loop is described in more detail in [23].

The cognitive functionalities of a CPHS, and thus of a CAV, are assumed to constitute a *cognitive layer* in their supervisory program(s). The repertoire of commands in the underlying programming system must be suited to allow for the programming of the desired functionalities of a cognitive nature.

Philosophy The baseline assumption leaves us short of a specification of the concrete cognitive abilities one would like cognitive CAVs to have. It is the aim of the postulates below to identify the philosophical principles for the 'shell' of cognitive abilities that are required for full autonomy of the vehicles.

Together with information about the environment and of other CAVs, the cognitive abilities of a CAV should empower its programs (or 'brain') to determine efficient routes and guarantee reliable and safe driving to the desired destination(s). Here 'efficiency' may be understood in any sense of the word: fast, economical, cleanly (or, environmentally friendly), and so on. Efficiency, reliability and safety are natural criteria for any system designed to serve us (like CAVs).

3 Autonomous vehicles and consciousness

Assuming the baseline for CAVs, the question arises how they must be facilitated so as to 'act and react awarely and intelligently with regard to their operation and their environment'. It is seen as crucial for autonomous driving at the highest level. In this section we digress on the requirement of *awareness* (or, *machine consciousness*) for driver-less vehicles and what can be achieved by it.

3.1 Postulate 1

CAVs use their sensory networks to sense their own state and the circumstances in their environment, and they use interactions with other CAVs around and the roadside 'intelligent traffic system' to plan and perform their mission. How can one enforce that CAVs are 'aware' of themselves and the environment, so they are facilitated to operate safely and reliably?

Safety and reliability are major issues in all CPHSs. In an earlier study [23] we argued that cognitive CPHSs should possess the qualities of human-like 'consciousness' in order to deal with these issue effectively. We subsequently advocated that, as a prerequisite for safety and reliability, all cognitive CPHSs should be designed to satisfy the requirements of a suitable machine version of this cognitive quality called *minimal machine consciousness*. By the baseline assumption it is only natural to require it for the design of CAVs as well, as expressed in the first of our postulates.

> **Postulate 1 (CAVs must possess Minimal Machine Consciousness)**
> *CAVs must be developed as cognitive CPHSs endowed with* minimal machine consciousness*, a prerequisite for their safe and reliable operation. Minimal machine consciousness processes both the car-dependent information from a CAV's own on-board units and car-independent information from other CAVs and the roadside traffic management system.*

The notion of minimal machine consciousness for CPHSs was proposed in [22, 23] and has until now not been made explicit in this form in the field of autonomous vehicles. Minimal machine consciousness intends to ensure 'machine awareness' of the system in which it is implemented. The feedback loops between the control unit and the sensory and motor units as imposed on the architecture of cognitive CAVs are drivers for the information that is needed for it.

In the remainder of this Section we elucidate the notion of minimal machine consciousness and what it brings for CAVs.

3.2 Minimal machine consciousness

The purpose of minimal machine consciousness (MMC) is to construct and maintain an abstract ('mental') model of a system's reality from which further knowledge and all of its meaningful actions can be derived [22, 23]. Specialized to CAVs, a minimally machine conscious system must possess the following properties:

(a) *self-knowledge*: it has complete knowledge of its current cognitive state as well as of the perceptual data produced by all its sensors, from other CAVs and from the roadside traffic system.

(b) *self-monitoring*: it has completely knowledge about the performance and status of its sensory and motor units during its operation. This includes the accuracy of the sensations and the reports from all of its sensory networks.
(c) *self-awareness* (or *self-reflection*): it behaves in a way that unambiguously reflects, resp. is determined by, its current cognitive state and the knowledge gained by its self-knowledge and self-monitoring abilities. With the help of introspection, self-distinction and change-detection (cf. Table 1), its control unit continually generates instructions for its motor units and the refreshment of its cognitive state.
(d) *self-informing*: it makes its cognitive state 'globally' available, i.e. to all modules of the system whenever changes of state occur (and to other CAVs and the roadside traffic system as needed).

Definition 1. *We say that a CAV is* minimal machine consciousness *(MMC) if and only if it satisfies the properties of* self-knowledge*,* self-monitoring*,* self-awareness *and* self-informing.

Note that minimal machine consciousness focuses on maintaining the cognitive state of a CAV and, by the later extension, of the CAVs around it. A more detailed description of the four principles of minimally machine conscious CAVs and their prevailing purpose is summarized in Table 1. (Cf. [22, 23].)

Principle	**Description**	**Purpose**
Self-knowledge	Informs the control unit what's going around based on the perception information from the on-board sensors and the roadside traffic system.	Informing the control unit on spatial and temporal location of the vehicle. Reporting.
Self-monitoring	Based on the feedback information from sensors and motors machine emotions are determined informing the control unit how well or if at all the sensors and motors work.	Confirms machine's certainty or error of its actions, enables repair of its own mistakes and damage detection
Self-awareness	Based on the current cognitive state and the information gained by self-knowledge and self-monitoring, and with the help of *introspection, self-distinction and change-detection,* the control unit computes instructions for its motor units *and* a new cognitive state. This state can also subsume any so-called *machine qualia states*, informing the unit about events requiring immediate attention and a remedy in the next step(s).	*Introspection* allows investigation of the own subjective past experience and emotions for decisioning purposes. *Self-distinction* enables the system to distinguish itself as an individual unit separate from other objects. *Change detection* allows awareness of changes in the outside world.
Self-informing	Global availability of the results of self-knowledge, self-monitoring, self-awareness and of the cognitive state for all modules.	Allows coordination and synchronization of all modules of the system.

Table 1. The four principles of minimal machine consciousness

In the later postulates we will set additional targets for a CAV to interpret, learn and help manage the traffic situations around it.

3.3 Reflection

Minimal machine consciousness is the basic cognitive mechanism of a self-driving vehicle. The presence of the mechanism is dictated by the requirement that, in order

for a CAV to behave purposefully as it is expected in its specification, its control unit must be able to construct and maintain a simplified, abstract model (representation) of reality from which further knowledge used for guiding the car can be derived. For this, the car must know its current cognitive state and have sufficient information from its on-board sensory and motor units and from the roadside traffic system on what is going around it as well as what is going 'inside' the vehicle, i.e. about the performance of its sensory and motor units. Only based on all these factors a CAV can subsequently elicit the behaviour leading to the fulfilment of its purpose and its understanding.

4 Autonomous Vehicles and Cooperation

We argued that a machine version of 'consciousness' is needed if we want CAVs to 'act and react awarely (and intelligently) with regard to their operation and the environment', but so far we emphasized their (inner) 'actions'. How can input from the *environment* ('the outside') help a CAV in gathering knowledge and becoming aware of its local and global context and in making informed decisions? In this section we digress on the role and requirements of a *smart roadside infrastructure* (SRI) and on the support it can provide to the vehicles in an application.

4.1 Postulate 2

CAVs that are minimally machine conscious (Postulate 1) will use all their connections and inputs, to construct and maintain a model of the (part of the) 'world' in which they must find their way. It is highly volatile information and thus, to support it, CAVs must be continually informed and updated about their local and global real-world context.

Although CAVs are 'connected' and can potentially inform each other without global control, CAVs in their vicinity and further abroad keep coming and going and they move in ever changing parts of the map. It is far more effective to delegate the task to a (global) roadside traffic system. This system must be designed to provide the information and feedback that will enable CAVs to draw the 'smart conclusions' for the mapping of their movements.

The design of minimally machine conscious CAVs that operate and cooperate towards achieving their set goals, should thus be seen in the broader perspective of designing and developing (large-scale) *smart roadside infrastructures* that comprise both the CAVs and (an abstraction of) their environment. In the paradigm of cognitive CAVs, SRIs should not only support, but also 'extend' the cognitive abilities of the vehicles in their scope and facilitate their cooperation by maximal information support, as expressed in the second of our postulates.

> **Postulate 2 (CAVs must be supported by Smart Roadside Infrastructures)** *The cognitive abilities of CAVs must be enhanced by a* smart roadside infrastructure (SRI)*, to enable effective and purposeful operation. The smart roadside infrastructure must be developed as an integral part of a CAV's cognitive control system. The car-independent information from the SRI complements the car-dependent information from a CAV's on-board sensory networks.*

It is implicit that there must be a steady interaction between a CAV and the SRI that represents its environment. It will enable a minimally machine conscious CAV to

exploit its cognitive abilities and the information from other CAVs and their environment in its AI-algorithms for effective route determination and for driving to a desired destination. One may include human-operated control centres in the SRI.

In the remainder of this Section we consider SRIs and how they may interact with CAVs in more detail.

4.2 Smart roadside infrastructure

SRIs support a multitude of functions and cognitive functionalities of autonomous vehicles. We outline some of them.

Cooperation We argued that, in order to cooperate effectively and purposefully, CAVs should cooperate with each other and the SRI. To this end, it is required that they possess the means for both *vehicle-to-vehicle* (V2V) and *vehicle-to-everything* (V2X) communication. The latter communication is typically supported by the global SRI, e.g. in the 'cloud'. It provides extra information support, adding to the inputs gained from the vehicle's on-board sensory networks [9]. In turn, a CAV can contribute its own findings to the SRI, for the benefit of other vehicles.

Roadside sensing The necessary connectivity and 'external' sensing capabilities for CAVs are to be embedded within the supporting SRI itself. This may be done via *smart roadside units* that support the wireless communication between the CAVs and any traffic-related objects in their surroundings. These units allow, for example, automatic communication with parking meters, parking garages, pedestrian crossings, street lights and other types of street mobiliary.

GPS signals, high definition maps and small *mobile roadside units* carried possibly by pedestrians (as part of their mobile phone, or in the form of small microprocessors), dogs, bicyclists, non-autonomous vehicles, and so on, can also be part of the SRI.

Global information The extra information from the SRI potentially relates to a larger neighbourhood than is covered by a vehicle's perception. Think of, for example, information on the state of the road ahead of the car, about nearby fuel or electric charging stations, about the weather, about accidents or obstacles ahead (including traffic congestions, demonstrations, road works, and so on) that cannot be recorded on the maps, about non-standard vehicles in the vicinity (garbage pickup truck, ambulance, trucks with oversized cargo, etc.), about vehicles 'round the corner' that cannot be spotted by the car's sensors, about humans that are about to cross the road, and so on. Some of this information may be redundant, but the fusion of V2X information with on-board information is a welcome strengthening of a vehicle's confidence in its perception.

World view As a result, with the help of smart roadside units and mobile roadside units, an autonomous vehicle can obtain a more accurate view of the world. This in turn facilitates to expand the scope of its machine consciousness and understanding to a segment covering a larger part of the real world than that mediated solely by the car's on-board sensors, thus helping to maximize safety [10].

5 Autonomous Vehicles and Machine Understanding

We postulated minimal machine consciousness as the requirement for CAVs to ensure that they can operate 'awarely' of their own state and of their environment. As CAVs

must 'act and react both awarely and *intelligently* with regard to their environment', these vehicles must also be able to understand (and act in) the traffic situations they may encounter, i.e. sense. In this section we digress on the requirement of *machine understanding* for driver-less vehicles and what it entails.

5.1 Postulate 3

The purpose of minimal machine consciousness was to enable a CAV to create a model of the perceived world in order that other qualities can evolve from it [22, 23]. In order for CAVs to act meaningfully and purposefully, such a model appears mandatory when it comes to understand and monitor their situation in the traffic. How can one ensure that CAVs 'understand' the situations they face and perhaps even 'learn' to deal with them?

Understanding (and learning) is a major issue in all CPHSs. We will argue below that all cognitive CAVs should be designed to satisfy the requirements of a suitable machine version of this cognitive quality called *machine understanding*, as expressed in the third of our postulates.

Postulate 3 (CAVs must possess Machine Understanding)
CAVs must be developed as cognitive CPHSs endowed with the power of machine understanding*, a prerequisite for understanding traffic situations and aware operation.*

In the remainder of this Section we define the notion of machine understanding and how it can be applied to CAVs. We will argue that, as expected, MMC is a necessary condition for CAVs to fully 'understand the situations on their mission'.

The idea of machine understanding appears to be a brand-new notion for the field of CAVs. In Section 6 we will discuss its possible realization with the help of a new online verification technique that guarantees safe trajectories in traffic situations, recently proposed by Pek *et al.* [13].

5.2 Machine Understanding: preliminaries

Prior to defining what we mean by machine understanding, we will take a closer look at the *epistemic domain* in which a CAV is expected to work, the mission of such a car, and the way of its operation.

We assume that CAVs operate on the streets in the real world. The epistemic domain of CAVs is thus a subset of the 'world' that is mediated to the vehicle's cognitive mechanisms, through its sensory networks. The vehicle's reactions to the perceived information are realized via changes of its cognitive state and via the execution of the respective commands issued to the its motor units.

Epistemic theories We also assume that there is an *epistemic theory* - a formal or less formal one - pertinent to the epistemic domain in question. The important feature of epistemic theories is that they allow one to describe *what* has to be done, rather than *how* it has to be done. This allows one to specify the acts of machine understanding for CAVs independent from an underlying computing machinery. (See e.g. [20, 21].)

In our case, an epistemic theory formally consists of the following three parts: a *descriptive* part, a *predictive* part, and an *executive* part. We briefly characterize these parts in turn.

A. The *descriptive part* describes the knowledge about the objects that can be perceived by the CAV's sensors at the beginning of a cognitive (SACA-)cycle of the underlying CPHS: their types (other cars, pedestrians, cyclist, and so on), their properties (position, size, vectors of movements), the relations among them (e.g., their distances) and the like.
B. The *predictive part* describes the rules for computing the future expected movements and positions of the objects described in the descriptive part of the theory, for a certain fixed period of time (of the order of fractions of seconds). Such predictions are based on the relations and measured dynamics of the objects traced in the descriptive parts, under the assumption that the *traffic rules* will be obeyed as they apply for these objects.
C. The *executive part* of the theory describes the rules for computing a safe trajectory of the car at hand for a given period of time and for issuing the respective instructions for the car's motor units. A safe trajectory is a trajectory that, in a given cognitive state of the system, allows the car to proceed safely from its current position towards its aim, taking into account the predicted movements of the objects and thus avoiding collisions with them, and the traffic rules, without endangering the other cars.

Missions For a given time moment, let us call the *traffic situation* at that moment be the ordered tuple of signals received from all sensory networks of the CAV, and likewise the *behaviour* of the vehicle's motor units as the ordered tuple of motor instructions sent to the its motor units at that moment of time.

Under the assumptions we stated above, the *mission* of a CAV can now described by a *relation* between the set of pairs of possible traffic situations and possible cognitive states on one hand, and the set of possible behaviours of the CAV on the other. This relation is defined with the help of the underlying epistemic theory in the following way: to each traffic situation and each cognitive state, the set of adequate behaviours is associated ('assigned') as determined by the executive part of this theory.

Obviously, this is a computable relation that can be computed by the vehicle's control unit.

5.3 Machine understanding: definition

We can now define the notion of machine understanding for driver-less vehicles as follows.

Definition 2. *Let E by an epistemic theory governing the behaviour of a vehicle C. Then we say that C, as a subject of machine understanding,* understands *traffic situations by means of theory E, if and only if for each traffic situation and each cognitive state, C executes the behaviour prescribed by the executive part of E. If this is the case, we say that C* fully understands *its mission w.r.t. theory E.*

Note that full (machine-)understanding of a vehicle's mission is an epistemic feat, rather than a computational property. In this sense it is similar to program correctness (cf. [8]). It cannot be "switched off" or "on". A system either has it, or not.

Comparing CAVs The extent of 'understanding' by a CAV very much depends on the underlying epistemic theory. In the simplest cases, as in our case, such a theory

covers but the most pragmatic and basic understanding of traffic situations. More developed theories can also offer explanations of the actions of a CAV, and epistemic theories covering large epistemic domains may call for a more general definition of understanding (cf. [19]).

Comparison of the abilities of self-driving vehicles based on their 'degree of understanding' clearly shows the futility of measuring such abilities by the number of 'disengagements', i.e. how often humans have to seize the wheel and take control of self-driving vehicles [17]. Namely, this number depends on the degree of a vehicle's understanding of its 'driving scenarios'. If the driving scenarios are incomparable, then the number of disengagements says nothing about the true self-driving abilities of the respective vehicles.

5.4 Machine understanding: necessity of MMC

'Understanding' commonly refers to the ability to build and work with a 'mental model' of some epistemic domain. This can be said of 'machine understanding' as well (cf. [14]). In our case, the domain consists of the knowledge generated by the self-awareness mechanisms of a CPHS and its processes, especially MMC, that constitutes such a mental model. In the case of CAVs, this knowledge refers to the objects within the perimeter of the vehicle's sensory networks, to their type, velocities, direction of their movement, and so on.

The following observation points to an important property of cognitive systems that fully understand their mission. It stresses the importance of MMC for machine understanding.

Proposition 1. *Minimal machine consciousness is a necessary condition for cognitive CAVs that fully understand their mission. However, it is not sufficient.*

Proof. (Sketch) Suppose a cognitive CAV fully understands its mission but is not MMC. In this case the CAV would not be fully self-knowledgeable, self-monitoring, self-aware or, indeed, not fully self-informing. Then situations may occur of which the perception characteristics are not completely, or not correctly, registered by the appropriate cognitive mechanisms of the CAV. Hence, due to the missing or incorrect information, the system will not be able to fully elicit the proper behaviour as required in its mission in these situations. Hence it does not fully understand its mission in these cases, which contradicts our supposition.

On the other hand, suppose MMC would be a sufficient condition for a cognitive CAV to fully understand its mission. Now consider, for example, any cognitive CAV that is MMC and, thus, self-aware of its environment. The information as it is produced by the general self-awareness mechanisms will be the same for all cognitive CAVs with this property: private cars, garbage pickup trucks, combatant vehicles, and so on. However, the difference in the missions of these vehicles is not only determined by their construction and the information produced by their self-aware mechanism, but also by the general context in which they operate and their understanding of it. Thus, being MMC alone is not sufficient, a contradiction again. □

It can happen that, although a system does not fully understand its mission, i.e., the entire set of situations in which it is assumed to operate, it *does* fully understand a

meaningful subset of such situations. This can be used for comparing the self-driving abilities of autonomous cars. In this case, the key notion is that of a driving scenario. A *driving scenario* is a set of *driving skills*, such as the ability to park, to change lanes on a highway, the ability to drive under poor light conditions, and so on. Now consider two different driving scenarios S_1 and S_2 such that $S_1 \subsetneq S_2$. Then a CAV that fully understands the situations in scenario S_2 would be able to fully understand the situations in scenario S_1 also, but not vice versa. This property can be used for the classification of autonomous vehicles (cf. [24]).

6 Cognitive CAVs and safe driving missions

In the baseline assumption of this report (Ch. 2) we advocated that CAVs are viewed as cognitive CPHSs, i.e. that they have the architectural and operational characteristics of such systems. The three postulates were developed as 'minimally required additional design objectives' for CAVs that can navigate and fulfil their purpose in a fully self-controlled manner. Have we reached our goal?

In order to answer this question we take a closer look at the key mechanisms of a (cognitive) CAV that satisfies the three postulates and how they cause the CAV to be aware of the changing environment, understand the situations with which it is confronted, and direct itself along a safe trajectory towards its destination. In Section 6.1 we sketch a *driving algorithm* based on these mechanisms that aims to guide a CAV in every step of the way. We argue that, under reasonable conditions, the algorithm indeed enables a CAV to safely reach its destination. In Section 6.2 we reflect on the further impact of the paradigm on safe missions.

6.1 A driving algorithm

Let $\mathcal{A}$ be a cognitive CAV that satisfies the baseline and the three postulates. We give a (high-level) description of an 'algorithm' that could be used by $\mathcal{A}$ to fulfil a driving mission. We follow the design paradigm of cognitive CPHSs and focus the description on the steps to be taken in the four consecutive phases of the operational cycle of $\mathcal{A}$, i.e. in its SACA loop (cf. Section 2).

Description In what follows we assume that at the beginning of the operational cycle, the control unit of $\mathcal{A}$ is in a given cognitive state called the *current cognitive state*. In this state also the current and final geographic positions of $\mathcal{A}$ are recorded. These positions are used by $\mathcal{A}$'s navigation system for determining its *current aim* – a trajectory on the 'map' from its current position towards its final destination.

By Postulate 1 we may also assume that $\mathcal{A}$ is minimally machine conscious, i.e. that it possesses the appropriate self-knowledge, self-monitoring, self-awareness and self-informing mechanisms. The operation of these mechanisms is described in the epistemic theory E governing the behaviour of $\mathcal{A}$ as a driver-less vehicle. We also assume that theory E captures the pragmatic aspects of 'understanding' that can be used in guiding the car's behaviour (cf. Postulate 3).

The driving algorithm for $\mathcal{A}$ as an autonomous vehicle that understands traffic situations by means of theory E, is depicted in Fig. 1.

Notes A remark concerning the categorization of objects in Phase 2 of the operational cycle is in order. Categorization is standardly performed by *neural nets*. These nets

Phase 1 – Sense: The data from all on-board sensors, navigation system of the car, and from V2X and V2V communication are retrieved with the help of self-knowledge and self-monitoring mechanisms. Especially, the current geographic position of the car is among these data.

Phase 2 – Analyze: Based on the previous data the self-distinction and change-detection mechanisms (parts of the self-aware mechanism) produce a mathematical model (a "mental model") of the current state of traffic that can be seen as a "formal specification" of the current traffic situation. Such a specification usually takes a form of a detailed on-board 3D map capturing all objects in the car's environment that must be taken into account when computing the next move function of the car. That is, each such object is categorized, whether it is a vehicle, of what type, a cyclist, a pedestrian, a person in a wheelchair, an animal, a lane marker, a traffic light, etc. Then position of each object w.r.t. the car must be stated, its vector of movement, and possibly its other characteristics. The respective new knowledge is generated by the *descriptive part* of theory E.

Phase 3 – Compute: Based on the mathematical model, current cognitive state, and introspection part of the self-aware mechanism, the value of the next move function is determined. This is done as follows using the *predictive part* of theory E. Starting from the mathematical model a so-called *motion prediction* — a detailed simulation of traffic by simplified dynamic models (cf. [13]) is performed. For each traffic participant a so-called reachability analysis is performed aiming at determination of its potential future positions at the end of some small time period (of the order of a fraction of a second) satisfying the car's current aim. The software also calculates potential emergency maneuvers in which the vehicle can prevent collisions by accelerating or breaking without threatening others. Then optimal trajectory for the car is selected. If all traffic participants adhere to traffic regulations such a trajectory must always exist [13]. A further advantage of this approach is that the respective computations are on-line verified.

Phase 4 – Act: Based on the computed optimal trajectory a new current aim is computed and the new cognitive state is determined. Corresponding new actions are determined using the *executive part* of theory E. New cognitive state and new actions are broadcast to all modules of the car.

Fig. 1. High-level description of the operational cycle of the driving algorithm of an autonomous car that cooperates and understands by means of epistemic theory E.

must be trained on labeled images. The images from the vehicle's cameras are labeled manually by human staff (cf.[3]). Because human labelling is error-prone, efforts to bypass the human input exist, but so far these efforts are not very successful. Humans draw boxes around the objects and describe their type (add annotations) very effectively. Eventually, having a sufficiently representative set of annotated images, a deep neural trained off-line on these data is subsequently used for on-line annotating the 'general' data from the cameras in real use by the vehicle.

Note that the resulting neural net plays the role of an *attentional mechanism* that selects from the real-world images delivered by the cameras, those objects of the underlying epistemic domain that are of interest for further processing (cf. [5]). Then the positions read from the radar data, ultrasound sensors, GPS coordinates, and high-definition maps are added to the annotated objects. The vectors (directions and speed) of moving objects are inferred from the Lidar data and computed from their previous positions. These vectors can further be validated or rectified by using the information from the surrounding cooperating cars. This is vital information that contributes to the precision of the obtained results.

Finally, based on the resulting information a mathematical model of the given situation is produced. Obviously, all future actions of the car depend on the accuracy of this model.

The effect of the given algorithm is summarized in the following observation. It confirms the effectiveness of the postulates.

Proposition 2. *Assuming that the epistemic theory E underlying $\mathcal{A}$'s operation provides the vehicle with full machine understanding of its mission, that all traffic participants adhere to the traffic regulations, and that all objects in the environment are recognized by the vehicle's sensory networks, then, starting from its current position and using the previous algorithm, $\mathcal{A}$ will eventually safely reach its final destination.*

Proof. (Sketch) If $\mathcal{A}$ fully understands it mission then, at the beginning of each operational cycle, it must receive enough information from its sensory networks about the epistemic domain in which it operates for its further steps. In particular, its self-awareness mechanism will produce enough knowledge for constructing a mathematical model of the vehicle's current traffic situation. Subsequently, using the algorithm and starting from the current cognitive state, thanks to the properties of the algorithm from [13] it will provably compute a safe trajectory for the car, reaching a new safe position at the end of the cycle in accordance with its current aim. That is, $\mathcal{A}$ understands its situation during the cycle at hand. By following an updated new current aim in each iteration of the operational cycle, $\mathcal{A}$ proceeds forward on the map to its final goal. After a finite number of iterations it will thus provably reach its final destination in a safe way. □

6.2 Reflections

Compared to the current approaches based entirely on the utilization of deep neural nets, the approach to *cognitive CAVs that cooperate and understand* according to our postulates, is likely to lead to safer missions for the following reasons.

a. First, a CAV's decisions are based on more information. Namely, the fusion of the vehicle's own perception with the information from a distributed SRI and the cooperating CAVs in Phase 1, is a means to provide vehicles like $\mathcal{A}$ with the maximal information available. The data obtained in this way is much more informative than the data a human driver can gain through his/her own senses. This is because of the sheer volume of such data, and because the data also contain information not accessible to human senses. In fact, utilization of the available information from multiple sources gives CAVs 'superhuman' perception abilities. The processing of this information by the cognitive mechanisms of the vehicles boosts their intelligence and results in safer driving (cf. [1]). This is achieved at the expense of more hardware, more communication and more computation.

b. Second, as shown in [13], the mathematical model constructed in Phase 2 of the operational cycle provably allows an online determination and verification of the optimal safe driving trajectory for arbitrary driving scenarios, even in cases when the classical approach based on neural nets fails. This failure inevitably happens in rare cases, in so-called *edge* situations. Edge cases concern the traffic situations not included in the training set of the respective neural net. Such cases occur because the set of training examples can never cover all traffic situations that may arise 'in practice' [11, 13]. Note that for the construction of the mathematical model, MMC

is indispensable: without this property, the mathematical model (in fact, the self-awareness property) would not hold in all situations. Therefore, testing for MMC is also of utmost importance. More about this subject can be found in [22, 23].

c. Finally, CAVs working in accordance with Proposition 2 are safer also in crash situations. Although in such situations the motion planing algorithm will not be able to come up with a safe trajectory, it will be able to determine a trajectory which will minimize the expected damage caused by unavoidable collision with some obstacle. To this end, also other cooperating vehicles can contribute, e.g. by freeing an escape route for a colliding car. In such situations the actions of the involved cars need not be aligned with the traffic rules. This kind of behaviour would not be possible without cooperation and when neural nets are used for route determination.

The algorithmic 'experiment' in this Section shows the extent of the postulates and their potential effect for improved safety and reliability.

7 Final remarks

In this report we set out to develop a new view on the design of (fully) autonomous vehicles from the perspective of the philosophy of computing, with the aim to identify high-level vistas that may lead to safer and more reliable vehicles. Starting from a powerful baseline assumption, we formulated three postulates for the design of CAVs which seem to bring us a long way towards turning them into 'cognitive machines' with the desired qualities. Some final remarks are in order at this point.

Interdependence and necessity of the postulates The three postulates we proposed are highly interdependent. Given the baseline that defines autonomous vehicles as cognitive cyber-physical human systems, the basic cognitive abilities are provided in Postulate 1 by the requirement of minimal machine consciousness. In order to maximize the reach of these abilities, input from the cooperating vehicles and the environment is needed and this is then provided by Postulate 2. Finally, Postulate 3 builds on Postulates 1 and 2 to complete the cognitive scope.

All three postulates are necessary, as shown in the algorithmic experiment in Section 6. Without Postulate 3 the cognitive abilities of CAVs would be deprived in strength and efficiency, while without Postulate 2 the vehicles would be forced to act under the poverty of stimuli in some situations. Without Postulate 1, there would be no analogue of 'conscious operation' in effect.

Note that the postulates at hand are technology independent – the only means needed for driving an autonomous vehicle is a computer-controlled mechanism exploiting cognition, artificial intelligence and communication. They also allow human input, facilitating basic ethical requirements. Last but not least, the postulates can easily be applied to all types of autonomous vehicles, whether they are road-borne, airborne or water-borne.

Paradigm Note that the objectives of the three postulates must all be developed hand in hand, from the very beginning of a CAV's design and development. One can not start with a classical car and then 'retrofit' it incrementally by additional sensory networks, advertising them as 'extras' for additional fees, although this is often done. This

is because cognitive CAVs require the specific architecture of cognitive cyber-physical human systems (cf. Section 2) with graded feedback from all sensor and motor units and mechanisms for minimal machine consciousness to emerge. Adding new sensors induces a change of the underlying self-awareness mechanism and hence, of the underlying epistemic theory that makes it understand its missions. This is a costly operation in terms of re-design and verification.

For the same reason, the cognitive abilities of a CAV can not be easily improved by including the information from the SRI: again, this requires a complete reworking of the vehicle's self-awareness mechanism.

SRI The previous observation has a profound effect on the methodology of designing autonomous vehicles and on the cost of their development. Practically, an SRI is effectively becoming part of the car's cognitive mechanism, and therefore their cognitive mechanisms must be designed and developed simultaneously, taking each other's existence into account from the beginning. This, of course, multiplies the price of the development of CAVs and complicates the development of autonomous transport enormously, since all of the necessary ingredients – the cars, the infrastructure, and the communication structure are in the hands of different stakeholders.

We see the present framework as a stepping stone for the further study of the formal theories and algorithms pertaining to the field of connected autonomous vehicles.

8 Conclusion

An analysis of the current mishaps with autonomous vehicles in practice unambiguously points to the main cause of their recent accidents: it is their insufficient understanding of a traffic situation at hand. It led us to formulate the following question:

> *Can one build the concept of safe and reliable CAVs around (machine versions of) the core notions from cognition like 'consciousness', 'cooperation', and 'understanding'?*

The approach developed in this report answers this question in the affirmative, applying viewpoints from the philosophy of computing and AI. Building on the baseline of viewing (connected) autonomous vehicles as cognitive CPHSs, we presented three postulates which together guarantee that the set goal can be reached, in principle. The essence of the design paradigm for CAVs can be summarized as follows.

B. *A CAV is a human-centric* cognitive cyber-physical human system *(CPHS) whose purpose is to transport people or goods with little or no human input, safely, reliably, and efficiently between two or more points on a given road map.*

P1 *Seeing CAVs as cognitive CPHSs, requires an architecture of the underlying system that allows them to be endowed with the mechanisms of* minimal machine consciousness*, to enhance their cognitive abilities.*

P2 *For route determination and driving to the desired destination safely, a CAV uses its cognitive abilities to* cooperate *with other CAVs* and *with the environment, notably through a* smart roadside infrastructure.

P3 *For dealing with the traffic situations on the way awarely and intelligently, a CAV must be endowed with the power of* machine understanding *so it can fulfil its missions safely and effectively.*

What the approach does, is offering a framework for how one might want to think about autonomous cars. Core notions are that of minimal machine consciousness, an integrated SRI, and machine understanding.

The framework is complex, but hopefully contributes to a versatile mindset for research on the design of CAVs. The trick leading swifter and safer to the final goal - a genuine fully autonomous human-centric CAV - is to see the problem of their development in a larger epistemic context, including both the vehicles and their environment. One should not consider cognitive CAVs merely as independent autonomous cybernetic devices, but as cognitive collectively collaborating devices endowed with features like minimal machine consciousness that provably lead to an understanding of their mission. Just this is the very thing that the new paradigm requires.

References

1. Andersen, H., Shen, X., Eng, Y. H., Rus, D., and Ang, M. H., Jr.: Connected Cooperative Control of Autonomous Vehicles During Unexpected Road Situations. ASME. *Mechanical Engineering* 139:12 (2017) S3-S7, https://doi.org/10.1115/1.2017-Dec-7
2. Baumberger, Ch., Beisbart, C., Brun G.: What is Understanding? An Overview of Recent Debates in Epistemology and Philosophy of Science. In: S.R. Grimm, C. Baumberger, S. Ammon (Eds), *Explaining Understanding: New Perspectives from Epistemology and Philosophy of Science*, Routledge, New York, 2017.
3. Bradshaw, T.: Self-driving cars prove to be labour-intensive for humans. *Financial Times*, July 9, 2017
4. Brandom, R.: Self-Driving Cars Are Headed Toward an AI Roadblock. *The Verge*, Jul 3, 2018
5. Chen, S., Jian, Z., Huang, Y., Chen, Y., Zhou, Z., Zheng, N.,: Autonomous driving: cognitive construction and situation understanding. *Science China Information Science* 62 (2019) article 81101. https://doi.org/10.1007/s11432-018-9850-9
6. Eliot, L.: Driverless Cars Must Be Self-Aware, A Crucial Missing Ingredient. *Medium*, May 8, 2019
7. He, J., *et al.*: Cooperative Connected Autonomous Vehicles (CAV): Research, Applications and Challenges. In: *2019 IEEE 27th International Conference on Network Protocols* (ICNP), 2019, pp. 1-6, doi: 10.1109/ICNP.2019.8888126.
8. Hoare, C.A.R.: Can Computers Understand Their Own Programs? Presentation in a special session on 'History and Philosophy of Computing'. In: *Computing with Foresight and Industry*, 15th Conference on Computability in Europe (CiE 2019), Durham, 2019
9. Holmes, F.: Smart infrastructure to provide extra support for autonomous vehicles. *Automotive World*, September 5, 2019, https://www.automotiveworld.com/articles/smart-infrastructure-to-provide-extra-support-for-autonomous-vehicles/
10. Jordan, M.I.: Artificial Intelligence - The Revolution Hasn't Happened Yet. In: *Medium*, April 19, 2018, https://medium.com/@mijordan3/artificial-intelligence-the-revolution-hasnt-happened-yet-5e1d5812e1e7
11. Koopman, P., Wagner, M.: Challenges in Autonomous Vehicle Testing and Validation. *SAE Int. J. Transportation Safety* 4:1 (2016) 15-24
12. Muller, H.A.: The Rise of Intelligent Cyber-Physical Systems. *Computer* 50:12 (2017) 7-9
13. Pek, C., Manzinger, S., Koschi, M., em et al.: Using online verification to prevent autonomous vehicles from causing accidents. *Nature Machine Intelligence* 2 (2020) 518-528
14. Sanz, R., Bermejo-Alonzo, J.: Consciousness and Understanding in Autonomous Systems. In: *Towards Conscious AI Systems* (TOCAIS 19), AAAI Spring Symposium, CEUR Workshop Proceedings Vol. 2287, paper 23, 2019, http://ceur-ws.org/Vol-2287/paper23.pdf
15. Sanz, R., Aguado, E.: Understanding and Machine Consciousness. *Journal of Artificial Intelligence and Consciousness* 07:02 (2020) 231-244
16. Seshia, S. A., Sadigh, D., Sastry, S. S.: Towards Verified Artificial Intelligence. In: *arXiv*:1606.08514v4 [cs.AI], 2020, https://arxiv.org/abs/1606.08514
17. Sumagaysay, L.: Self-driving companies: Don't measure us by 'disengagements'. *Protocol*, February 26, 2020, https://www.protocol.com/measuring-waymo-self-driving-readiness

18. Sowe, S.K., Simon, E., Zettsu, K., de Vaulx, F.F., Bojanova, I.: Cyber-Physical Human Systems: Putting People in the Loop. *IT Professional*, 18:1 (2016) 10-13
19. Thórisson, K.R., Kremelberg, D.: Do Machines Understand? A Short Review of Understanding & Common Sense in Artificial Intelligence In: *10th Int Conference on Artificial General Intelligence* (AGI-17), AGI-17 Workshop: *Understanding Understanding*, Melbourne, Australia, Aug. 18, 2017
20. Wiedermann, J., van Leeuwen, J., What is computation: An epistemic approach, In: G.F. Italiano *et al.* (Eds.), *SOFSEM 2015: Theory and Practice of Computer Science*, Proc. 41st Int Conference on Current Trends in Theory and Practice of Computer Science, Lecture Notes in Computer Science, Vol. 8939, Springer, 2015, pp. 1-13
21. Wiedermann J., van Leeuwen J.: Epistemic Computation and Artificial Intelligence. In: Müller, V.C. (Ed.), *Philosophy and Theory of Artificial Intelligence 2017* (PT-AI 2017). Studies in Applied Philosophy, Epistemology and Rational Ethics, Vol. 44, Springer, Cham, 2018, https://doi.org/10.1007/978-3-319-96448-5_22
22. Wiedermann, J., van Leeuwen, J.: Finite State Machines with Feedback: An Architecture Supporting Minimal Machine Consciousness. In: F. Manea *et al.* (Eds), *Computing with Foresight and Industry*, 15th Conference on Computability in Europe (CiE 2019), Lecture Notes in Computer Science, Vol. 11558, Springer, 2019, pp. 286-297
23. Wiedermann, J., van Leeuwen, J.: Towards Minimally Conscious Finite-State Controlled Cyber-Physical Systems: A Manifesto. In: T. Bureš *et al.* (Eds.), *SOFSEM 2021: Theory and Practice of Computer Science*, Proc. 47th Int Conference on Current Trends in Theory and Practice of Computer Science, Lecture Notes in Computer Science, Vol 12607, Springer-Verlag, 2021, pp. 43-55.
24. Wikipedia: *Self-driving car*, July 2020, https://en.wikipedia.org/wiki/Self-driving_car
25. Wikipedia:*Understanding*, July 2020, https://en.wikipedia.org/wiki/Understanding#Shallow_and_deep