Proceedings of the 33rd European Safety and Reliability Conference (ESREL 2023)
Edited by Mário P. Brito, Terje Aven, Piero Baraldi, Marko Čepin and Enrico Zio

doi: 10.3850/978-981-18-8071-1_P423-cd

On the Foundation of Autonomous Mobility: Establishing Fundamental Principles for a Digital Driver

Bård Myhre

SINTEF Digital, Norway, E-mail: bard.myhre@sintef.no

Are Hellandsvik

SINTEF Digital, Norway, E-mail: are.hellandsvik@sintef.no

Stig Petersen

SINTEF Digital, Norway, E-mail: stig.petersen@sintef.no

Autonomous mobility is by many regarded as the ultimate vision for future green transportation, and driverless operations are therefore pursued for several transportation modes, such as road, rail and water. The scientific and technological communities within each of these domains do however employ different definitions of autonomy or automation. And while these definitions to a certain extent describe the functions of the vehicle, train or vessel, they do not provide any assistance or direction when it comes to designing autonomous systems. This paper aims to address these shortcomings by suggesting an alternative way of approaching autonomous systems. After presenting a set of overall requirements for a common framework for autonomous transportation, the paper introduces the concept of a "digital driver" as a new basis for autonomous mobility. The paper then suggests five fundamental principles for a digital driver, and describes how these may contribute in the analysis and design of autonomous mobility systems.

Keywords: Autonomy, digital driver, responsibility, liability, human-machine interaction, remote-control, framework.

1. Introduction

Autonomous mobility is by many regarded as the ultimate vision for future green transportation, and driverless operations are therefore pursued for several transportation modes, such as road, rail and water. The scientific and technological communities within each of these domains do however employ different definitions of autonomy or automation: Road transportation mainly adheres to the six "SAE Levels of Driving Automation" (SAE International 2018), rail automation is characterized using five "Grades of Automation" (IEC 62290-1:2014), while the maritime domain seems to narrow down to five "degrees of automation" (DNV GL 2018; Bureau Veritas 2019). While these definitions to a certain extent describe the *functions* of the vehicle, train or vessel, they do not provide any assistance or direction when it comes to *designing* autonomous systems. Furthermore, they do not touch upon aspects such as hand-over between humans and automation, nor do they provide a conceptual framework on how to handle the shift between remote control and autonomous control.

This paper aims to address these shortcomings by suggesting an alternative way of approaching autonomous systems, based on the concept of a "digital driver". It is our belief that shifting the focus from a *vehicle's capabilities* to a *driver's responsibilities* will open for new and constructive ways of assessing, discussing and designing both autonomous and remote-controlled systems. We also believe that this mindset will allow for a more natural interaction between humans and automation, including how to properly handle requests for humans to intervene. Finally, we think that establishing the *digital driver* as the foundation of autonomous mobility will provide a useful framework when developing new regulations and when it comes to handling liability in autonomous operations.

To give an overview of both the identified challenges and suggested solution, the paper is structured as follows: In section 2, we give an introduction to existing definitions of autonomy within road, rail and maritime, and argue why these definitions do not provide the adequate foundation for developing autonomous systems. Section 3 follows up with elaborating on which qualities a suitable framework for autonomous mobility *should have*, inviting to a high-level contemplation on why the selection of framework matters and how unsuitable and conflicting definitions hinder technological development. Section 4 then introduces what we suggest as a new foundation for autonomous mobility, by proposing *five fundamental principles for a digital driver*. In section 5, the suggested principles are applied to a basic case, illustrating how the concept of a digital driver enables us to take new perspectives on un-manned mobility. Section 6 gives a brief evaluation of how the suggested principles fulfil the overall need for a common framework, while section 7 considers some of the practical implications of the presented principles. Our tour then slows down with section 8 pointing forward to remaining challenges, before section 9 concludes the paper.

2. Existing Definitions of Autonomy – Shortcomings and Challenges

In their report on "Human and Computer Control of Undersea Teleoperators", Thomas B. Sheridan and William L. Verplanck (1978) suggest ten different ways – called *levels of automation* - that a human operator could interact with a computer that does the actual implementation of a specific task. The levels of automation suggested by Sheridan and Verplanck range from "human does the whole job up to the point of turning it over to the computer to implement" (level 1) to "computer does the whole job if it decides it should be done, and if so tells human, if it decides he should be told" (level 10).

Although the suggestions by Sheridan and Verplanck are not in direct use in any transportation domain, their approach with defining specific *levels of automation* have gained foothold within both road, maritime and rail transportation. In the following sections, we will briefly describe how the three domains approach automation.

2.1. *Automation within road transportation*

Within road transportation, SAE International has developed a *recommended practice*, denoted SAE J3016. SAE J3016 defines six levels of driving automation, with each level describing the roles of a *human user* and of a *driving automation system*. At level 0 (No Driving Automation), the human user performs the entire driving operation. At the other extreme, level 5 (Full Driving Automation) describes a driving automation system that can perform a driving operation sustainedly and unconditionally with no operational limits.

While one may argue that achieving level 5 – with driving under *all* conditions – can only be regarded an ambition and not a reachable goal, we consider the main challenge with the SAE J3016 to be more mundane. Within driving level 3 (Conditional Driving Automation), the automation will perform sustainedly under many conditions, but the driver is still required to be ready for driving during *fallback*. This means that the automated system at any time – although possibly with a warning – may leave the control of the vehicle to the human driver. Although there has been some critique of the SAE conditional driving automation definition (Inagaki and Sheridan, 2018), this potential critical flaw is still not addressed by the recommended practice.

2.2. *Automation within maritime transportation*

Whereas road transportation seems to have settled for the principles suggested by SAE J3016, the maritime domain is somewhat more diverse. To exemplify, a scoping exercise initiated by IMO in 2017 (IMO 2021) has defined the following four *degrees of autonomy*:

- Degree One: Ship with automated process and decision support
- Degree Two: Remotely controlled ship with seafarers on board
- Degree Three: Remotely controlled ship without seafarers on board
- Degree Four: Fully autonomous ship

Based on this categorization, DNV GL (DNV GL 2018) and Bureau Veritas (Bureau Veritas

2019) each have released documents defining five almost similar *degrees of automation* (here as presented by Bureau Veritas):

- A0 – Human operated: All operations are under human control
- A1 – Human directed: System suggests actions, but human makes decisions (and actions)
- A2 – Human delegated: System invokes functions, human must confirm
- A3 – Human supervised: System invokes functions, human is informed of decisions
- A4 – Full automation: System invokes functions, human is informed in case of emergency

Finally, Bureau Veritas also has defined eight *degrees of control* (Bureau Veritas 2019), covering both direct control (DC) from onboard personnel as well as remote control (RC) from personnel in a Remote Control Center (RCC):

- DC0: No direct control
- DC1: Available direct control
- DC2: Discontinuous direct control
- DC3: Full direct control
- RC0: No remote control
- RC1: Available remote control
- RC2: Discontinuous remote control
- RC3: Full remote control

With regard to the abovementioned degrees of *autonomy*, *automation* and *control*, we would like to point out some observations. First, the four *degrees of autonomy* may indicate that remote control is an intermediate step towards full autonomy. While this may be true in some cases, one could also look at remote control and autonomous control as two different modes of control, potentially with different capabilities for different types of operations. On this background, one could argue that this categorization is misleading. Second, regarding the five *degrees of automation*, we would like to point the attention towards degree "A2 – Human delegated", where a human must confirm functions that are to be invoked by the system. At first sight, this may seem like a sound way to ensure that a competent human is always in charge. Several studies, as reported by Matthews et al. (2019), have however illuminated the challenge of the *passive fatigue* that is produced by monotonous work. Although one could argue that this is a matter of finding ways to counteract the passive fatigue, we think one should also consider a more radical approach where such a work mode is not an option. Third, when it comes to the eight *degrees of control*, these represent "the degree of availability of human operating the ship aboard (crew) or remotely outside the ship from a remote control centre (operators)" (Bureau Veritas 2019). This may seem like a practical categorization, indicating what modes of control are available. It does however not indicate who *has* control, nor does it cover autonomous control. Therefore, we propose that this classification does not necessarily provide any significant help when it comes to describing and understanding a vessel's capabilities.

As a concluding remark to autonomy in the maritime domain, we think it may be challenging to have three different types of classifications without describing their purpose and how they are expected to be used. This leads us to an idea that any classification should *fulfil specific needs*, so that the technical community can evaluate its usefulness.

2.3. *Automation within rail transportation*

To our knowledge, rail is the only mode of transportation where driverless mobility is already offered as a commercial service. Within rail transportation, the standard IEC 62290-1:2014 defines five *grades of automation* (GOA), ranging from GOA0 to GOA4:

- GOA0: On-sight train operation
- GOA1: Non-automated train operation
- GOA2: Semi-automated train operation
- GOA3: Driverless train operation
- GOA4: Unattended train operation

In contrast to similar definitions within other modes of transportation, the *grades of automation* in rail strictly define what capabilities are required by the system (that controls the operation) with respect to the basic functions of train operations: GOA1 requires the system to ensure safe separation of trains, GOA2 requires the system to control acceleration and braking, GOA3 requires the system to prevent collisions with obstacles and persons, while GOA4 requires the system to supervise passenger transfer, take trains in and out of operation, and supervise the status of the trains.

What is particularly interesting with the grades of automation as given by IEC62290-1:2014, is that they all cover one or more specific basic functions that are related to the safe operation of the train. This means that each GOA in practise represents a *collection of specific requirements*. The IEC62290 grades of automation therefore serve as a well-defined framework in the design and understanding of the overall rail service, by directly influencing requirements on operation, operational facilities, rolling stock and staff.

3. Requirements for a Framework for Autonomous Mobility

In the previous section, we have presented how autonomous mobility is defined within road, maritime and rail transportation. Furthermore, we have suggested that a classification or definition of autonomy should fulfil specific needs and be useful in the design and analysis of the system. Also, we have stated that rail is the only domain that so far has a definition that supports the design of autonomous solutions. Based on this, we think it is timely to investigate what a framework for autonomous mobility should ideally look like.

3.1. *The purpose of a common framework*

The main purpose of a common framework for autonomous mobility should be to ensure a common understanding between actors that want to specify, develop, deploy, operate and maintain solutions for autonomous mobility. We further propose that a successful framework should be applicable for all modes of transportation, serve as a tool for communication between different actors, provide guidance and a common terminology, and be regarded as useful and relevant throughout the entire lifecycle of the solution.

3.2. *Specific requirements for a framework for autonomous mobility*

An overall framework for autonomous mobility would necessarily need to include all types of control, including human control and remote-control. It should also cover all variants of public and private areas, international transportation, and emergency situations.

Based on this, we suggest that a successful framework should at least:

- Be applicable and relevant for all modes of transportation
- Cover human control, autonomous control and remote control
- Cover manned, unmanned and partially unmanned operations
- Cover operations in both closed, private and public areas
- Support handover between different modes of control
- Support manual operations, decision support, semi-automated operations and fully automated operations
- Support taking back control through forced hand-over
- Support emergency situations as an integral part of the framework
- Take into account that not all control modes may be allowed under every environmental condition and geographical location
- Allow for different countries having different rules and legislation when it comes to automatic and remote operations
- Allow for certification by national and international transportation authorities

Although there may be more requirements, the ones above will expectedly work well as a starting point. In any case, we think that the idea of discussing *requirements for a common framework for autonomous mobility* is an important contribution to field of autonomous transportation.

4. Fundamental Principles for a Digital Driver

In order to establish a common and generic framework for autonomous mobility, we suggest introducing the concept of a *digital driver*. Our *digital driver* is an equivalent to the traditional *human driver*, who operates a *mechanical vehicle* through a defined interface. For both digital and human drivers, the vehicle in question can be of any type, operating on land, on water, under water and in the air.

By basing our understanding for autonomous mobility on the digital driver, we believe that we will be able to create a common framework for all types of transportation. We do however need to establish some basic principles regarding the digital driver, answering the following questions: What is a digital driver, who is responsible for its actions, and how is this responsibility transferred to and from the digital driver?

4.1.The r*esponsibility and liability of a digital driver*

In a previous paper (Myhre et al. 2019) we have suggested the idea that autonomy should be defined in terms of *responsibility* instead of *capabilities*. In this paper we aim to combine this idea with the concept of a digital driver, and we start out by suggesting that the defining characteristic of a *digital driver* is that it can assume responsibility for the control of a vehicle.

> **Principle 1.** A digital driver is a digital entity that may assume responsibility for the control of a vehicle.

This principle first states that a *digital driver* is a *digital entity*, which may at first glance seem obvious. It does however indicate the importance of knowing the boundaries of the digital driver, and that the *entity in question* is unambiguously defined. This means that the digital entity may be a *distributed* control logic in the cloud, but it may also be a *local on-vehicle* control logic with no external connections. What is crucial, is that there is no doubt related to where the entity ends. Next, the principle states that the digital driver may *assume responsibility for the control of the vehicle*. Here, the two keywords are *responsibility* and *control*. If a digital driver has the responsibility for the control, it means that it cannot rely on any others for the control of the vehicle. It also means that the digital driver has the full responsibility for maintaining a safe operation, under *all conditions*. This specifically contradicts the SAE J3016 driving level 3 (Conditional Driving Automation), which suggests a human fallback in case of problems. It is however important to understand that our definition does *not* imply that a digital driver must be able to control the vehicle under all conditions. It simply means that the digital driver must be able to assess the situation and avoid entering situations that it cannot handle. Just like any human driver continuously does (or at least should continuously do).

Our next principle addresses who has the legal liability for the actions of a digital driver. Alawadhi et al. (2020) affirm that there are yet no general rules on liability for autonomous vehicles, and we therefore take the liberty of placing the responsibility for a digital driver with its manufacturer.

> **Principle 2.** Any action caused by a digital driver is the legal responsibility (liability) of the manufacturer of the digital driver.

This principle has several consequences. First, it means that the digital driver is not a legal subject by itself. Second, it implies that no human driver should be considered liable for *any action* caused by a digital driver, even in cases where the human driver is the one who initiated the digital driver. Third, it states that the manufacturer may not delegate the liability to any other party. Even though delegation of liability can possibly be achieved by redefining the manufacturer role, it nonetheless ensures that the liability is held by a specific party.

It should be noted that allowing the liability of a digital driver to be held by its manufacturer will contradict the existing convention that a responsible person must at all times be able to control their vehicle (Bartolini, 2017). Therefore, this principle will probably require either an adaption of national laws, updating the Vienna Convention on Road Traffic of 1968, or both.

If the liability for a digital driver is to be held by the manufacturer, it becomes crucial to ensure that the digital driver is not considered responsible (or liable) without the manufacturer being "comfortable" with the responsibility. This leads us to the next challenge, namely addressing how responsibility is transferred to a digital driver.

4.2. *Transfer of responsibility to a digital driver*

When human drivers are to be responsible for a vehicle, they are expected to make considerations on whether the vehicle is safe to operate within the current and expected conditions. And if they find that the situation is not safe, they are expected not to operate the vehicle until they deem it safe. We suggest that the same philosophy should apply to digital drivers, implicating that a digital driver cannot be *forced* to assume responsibility for a vehicle. Furthermore, we suggest that a digital vehicle cannot assume responsibility for a vehicle without being offered the responsibility by a human. This leads us to the next two principles, that cover *offering* and *accepting* (or *refusing*) responsibility. We start with looking at the offering part.

Principle 3. A digital driver may be offered responsibility for the control of a vehicle by a human or another digital driver.

This means that a human (possibly a human *driver*, but not necessarily) or another digital driver may offer a digital driver responsibility for a vehicle, and that this is the only way a digital driver may receive responsibility for a vehicle. Note that as no digital driver can assume responsibility out of nowhere, this further means that all "chains of responsibility" must be initiated by a human.

We then turn to the other side of the offer, by looking at what a digital driver does when it is offered responsibility for a vehicle.

Principle 4. A digital driver that is offered responsibility for the control of a vehicle may accept or refuse the offer.

This principle states – in all its simplicity – one thing, and one thing only: That the digital driver must be able to make its own decision on whether or not to accept the responsibility for a vehicle. Although this is exactly what is expected from human drivers, we are not aware that this has been considered by others as a fundamental characteristic of autonomous systems. The consequence of this principle is that one should never expect a digital driver to be willing to obey. Conversely, if a digital driver has first accepted responsibility, it will keep that responsibility until the operation is completed. Another implication is that *defining the scope of the operation* is of utter importance when it comes autonomous operations, as this may have direct implications on whether a digital driver will accept or refuse an offer for responsibility.

4.3. *Retracting responsibility from a digital driver*

We have already stated that digital drivers can be offered responsibility, and that they may choose to accept or refuse that offer. The last principle describes how responsibility may be withdrawn from a digital driver in a forced transfer of responsibility.

Principle 5. A human or digital driver that has transferred the responsibility to another driver, can at any time retract the responsibility for the control of the vehicle.

This opens for *withdrawing* a responsibility that has previously been transferred, but only from the drivers that have held (and offered) responsibility for the same operation. With this principle, we also implicitly create the concept of rank (or hierarchy) with regard to which drivers that can (and cannot) retract responsibility from others. Note that once responsibility is retracted, it stays with the digital or human driver that retracted it. If that driver wants to transfer the responsibility back to the former human or digital driver, it once more has to be *offered* according to Principle 3. And *accepted* or *refused* according to Principle 4.

5. Applying the Principles

To illustrate how the principles can be applied in practise, we will briefly show how a simple road-based case can be interpreted based on the suggested framework.

5.1. *Case: Digital car driver accepts responsibility for parts of trip*

This case covers a car trip through waypoints A-B-C-D, where a *Human driver* is certified for driving the whole trip (A-B-C-D). The car also has a *Digital driver* that is certified for driving the car from waypoint B to C.

<u>Scope of operation</u>

- Drive car through waypoints A-B-C-D

<u>Actors</u>

- *Human driver*, certified for waypoints A-D
- *Digital driver*, certified for waypoints B-C

<u>Operational steps</u>

Step 1: *Human driver* drives car from A to B.

Step 2: Upon reaching waypoint B, *Digital driver* informs *Human driver* that it now can assume responsibility until waypoint C (according to Principle 1).

Step 3: *Human driver* offers responsibility to *Digital driver* (according to Principle 3).

Step 4: *Digital driver* accepts responsibility (according to Principle 4) and takes control of car. The manufacturer of the *Digital driver* has now all liability (according to Principle 2). *Human driver* has no responsibility for operation but may retract control (according to Principle 5) at will.

Step 5: When approaching waypoint C, *Digital driver* informs *Human driver* that it soon will not be certified to operate car.

Step 6 – alternative A: *Human driver* retracts control and responsibility before reaching waypoint C (according to Principle 5), and operates car until waypoint D.

Step 6 – alternative B: *Human driver* does not retract control (according to Principle 5), and *Digital driver* stops the car safely before waypoint C (according to Principle 1). *Human driver* resumes control at will and drives to waypoint D.

5.2. *Considerations on the presented case*

Although the presented case is a rather basic one, it provides us with some useful observations regarding the applicability of the suggested framework.

First, we see that the *Digital driver* after a specific point informs the *Human driver* that it can now assume responsibility. This means that the *Digital driver* in some way must be certified for a specific distance or area. Most probably, this will require the road authorities to issue a licence to the *Digital driver* for the specific distance.

Second, we observe that although the *Digital driver* informs that it can assume responsibility, there is still a formal offer of responsibility, and a formal acceptance, before the actual transfer of responsibility. This ensures that the *Digital driver* may reject the transfer (and the corresponding responsibility) if the *Human driver* has not sufficient control over the situation when offering responsibility.

Third, when the *Digital driver* has accepted responsibility, it maintains full responsibility until it either stops the car or the *Human driver* retracts the responsibility. There is no "partial responsibility", like the SAE J3016 driving level 3 - Conditional Driving Automation), only full responsibility or no responsibility.

Finally, we see that although the *Digital driver* is not able to drive in all areas under all conditions, it can still serve a purpose. What is important, is that the *Digital driver* knows its own limitations, and that it only accepts responsibility when it can fulfil its obligations. This may resemble what is required by ordinary human drivers, who are not expected to be *perfect*, but *qualified* for the purpose.

6. Evaluating the Principles

A thorough evaluation of the suggested principles in section 5 against the requirements in section 3 is not feasible within the format of this paper, but some aspects deserve being commented:

- The suggested framework should be relevant for any mode of transportation that can employ the concepts of human and digital drivers.
- All types of controls (human, autonomous, remote, manned, unmanned and partially unmanned) may be covered by the framework.
- Both closed, private and public areas fit into the framework.
- Handover between different modes of control is supported.
- For semi-automated operations there will be a need to define who holds responsibility.
- Forced hand-over is handled as "retraction of responsibility".
- Emergency situations must be handled by the current driver. This means that any driver must be prepared for handling emergency situations on their own, and not rely on external fallback solutions.
- Environmental and geographical limitations will be considered by each driver before accepting responsibility, and the expected scope of the operation must be taken into consideration.
- Certification of digital drivers should be handled by national transportation authorities the same way as for human drivers.

7. Practical Implications of the Principles

The practical implications of the suggested principles are expected to be found across the entire life-cycle of autonomous and remote-controlled transportation solutions, starting with the specification phase. Here, the principles may contribute in analysing which digital and human drivers are needed for a specific operation, their interactions, and their required capabilities. Furthermore, the principles may assist in considering how to approach emergency situations, environmental and geographical limitation, as well as the need for certification. They may also serve as guidelines when it comes to sorting out responsibilities between the different actors that constitute the overall transportation system.

8. Future Work

Having suggested a new basis for approaching autonomous mobility, we expect that there are challenges to be addressed on many levels. First, the suggested requirements for the framework (in section 3) deserve being revised for completeness and validity. Also, there may be other requirements that we have not yet identified, and that should have been included. Second, we do not expect that the principles that we have put forward (in section 4) are set in stone, and we welcome any contributions to develop these further into something that can be of use for the transportation community. Third, applying the principles to real cases will expectedly be the ultimate test, showing what value the principles will have when it comes to dismantling complex cases of autonomy, remote-control and manual control, as well as interactions between various types of digital and human drivers.

9. Conclusions

In this paper, we have argued why existing definitions of autonomy have substantial shortcomings, and why there is need for a common framework for autonomous mobility across all modes of transportation. Furthermore, we have presented a set of requirements that may provide a valuable starting point when developing such a framework. Finally, we have proposed five principles for a digital driver, that potentially may serve as a common foundation for autonomous mobility across all domains.

Acknowledgement

The results presented in this paper are based on previously unpublished work in the Research Council of Norway's Sacomas project under grant agreement No. 269510, and on subsequent discussions with Norwegian and international actors pursuing the field of autonomous transportation.

References

Alawadhi, M., Almazrouie, J., Kamil, M., Khalil, K.A. (2020). Review and analysis of the importance of autonomous vehicles liability: a systematic literature review. *Int J Syst Assur Eng Manag. 11(6),* 1227–1249

Bartolini C., Tettamanti T., Varga, I. (2017). Critical features of autonomous road transport from the perspective of technological regulation and law. *Transportation Research Procedia 27 (2017),* 791–798

Bureau Veritas (2019). Guidance Note NI 641: Guidelines for Autonomous Shipping

DNV GL (2018). Class guideline DNVGL-CG-0264: Autonomous and remotely operated ships

IEC 62290-1:2014. Railway applications - Urban guided transport management and command/control systems - Part 1: System principles and fundamental concepts

IMO (2021). Outcome of the regulatory scoping exercise for the use of maritime autonomous surface ships (MASS).

Matthews, G., Neubauer, C., Saxby, D.J., Wohleber R.W., Lin J. (2019). Dangerous intersections? A review of studies of fatigue and distraction in the automated vehicle. *Accident Analysis & Prevention, Vol. 126*, 85-94.

Myhre, B., Hellandsvik, A., Petersen, S. (2019). A responsibility-centered approach to defining levels of automation. *J. Phys.: Conf. Ser. 1357 012027*

SAE International (2018). Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles (Surface Vehicle Recommended Practice: Superseding J3016 Sept 2016)