IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 9 November 2022, accepted 4 January 2023, date of publication 13 January 2023, date of current version 19 January 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3236814

APPLIED RESEARCH

A Contextual Multimodal System for Increasing Situation Awareness and Takeover Quality in Conditionally Automated Driving

MARINE CAPALLERA[1], QUENTIN METEIER[1], EMMANUEL DE SALIS[2], MARINO WIDMER[3], LEONARDO ANGELINI[1,4], STEFANO CARRINO[2], ANDREAS SONDEREGGER[5], OMAR ABOU KHALED[1], AND ELENA MUGELLINI[1]

[1]HumanTech Institute, University of Applied Sciences and Arts of Western Switzerland (HES-SO), 1700 Fribourg, Switzerland
[2]He-Arc, University of Applied Sciences of Western Switzerland (HES-SO), 2000 Saint-Imier, Switzerland
[3]Department of Informatics, University of Fribourg, 1700 Fribourg, Switzerland
[4]School of Management Fribourg, University of Applied Sciences and Arts of Western Switzerland (HES-SO), 2800 Fribourg, Switzerland
[5]Business School, Institute for New Work, Bern University of Applied Sciences, 3008 Bern, Switzerland

Corresponding author: Marine Capallera (marine.capallera@hes-so.ch)

This work was supported by the Hasler Foundation (Switzerland) in the framework of the AdVitam Project.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethics Committee of the University of Fribourg under Approval No. 445.

ABSTRACT Advanced driver assistances are becoming increasingly common in commercial cars, not only to assist but also to free drivers from manual driving whenever possible. Soon, drivers should be allowed to engage in non-driving-related tasks. The fact that responsibility for driving is shifting from humans to machines must be considered in the development of these assistances in order to guarantee safety and trust. In this article, we introduce AdVitam (for Advanced Driver-Vehicle Interaction to Make future driving safer), an autonomous system aiming at maintaining driver's situation awareness and optimizing takeover quality during conditionally automated driving. The information conveyed to drivers is dynamically adapted to achieve these goals, depending on the driving environment and the driver's physiological state. This system consists of three connected modules. The first module (***Driver State***) predicts the driver's state with machine learning and physiological signals as inputs. The second module (***Supervision***) uses different interfaces (a haptic seat, a personal device, and ambient lights) to maintain the drivers' situation awareness during the autonomous driving phases. The third module (***Intervention***) is a machine learning model that chooses the most appropriate combination among haptic, auditory, and visual modalities to request the driver to take over control and thus optimize takeover quality. To evaluate the system and each module independently, a preliminary user study with 35 drivers was conducted in a fixed-base driving simulator. All participants drove in two different environments (rural and urban). In addition, the activation of the ***Supervision*** and ***Intervention*** modules were manipulated as two between-subject factors. Results show that conveying information on the driving environment status through multimodal interfaces increases drivers' situation awareness (i.e., better identification of potential problems in the environment) and trust in the automated vehicle. However, the system does not show positive outcomes on takeover quality. Besides, the ***Driver State*** module provided consistent predictions with the experimental manipulation. The system proposed in this paper could lead to better acceptance and safety when conditionally automated vehicles will be released by increasing drivers' trust during phases of automated driving.

INDEX TERMS Automated driving, driver's state, human–vehicle interaction, machine learning, multi-modal interaction, situation awareness, takeover quality, trust.

The associate editor coordinating the review of this manuscript and approving it for publication was Vicente Alarcon-Aquino.

I. INTRODUCTION

A. CONTEXT

Even if the number of road accidents is decreasing year after year in many countries, this remains a major cause of death in the world [1]. Usually, drivers are responsible for accidents causing material and physical damage, often due to an error or inattention [2]. To alleviate this problem, car manufacturers have been working on vehicle automation, precisely to support the driver. To provide a common basis for the scientific and industrial community, the Society of Automotive Engineers (SAE) has introduced a taxonomy of terms associated with vehicle automation. In this article, we defined LX-SAE as the level of automation corresponding to the level X of automation in the SAE taxonomy. In 2022, partially automated vehicles (L2-SAE) are on the road, equipped with advanced driver assistance systems (ADAS) to assist the driver in longitudinal (e.g., speed, distance to the vehicle in front) and lateral (e.g., steering wheel control and lane positioning) control of the vehicle. If we follow the taxonomy, the world should see the emergence of conditionally automated vehicles (L3-SAE) in the near future, before vehicles are fully autonomous (L5-SAE). Currently, L3-SAE automated vehicles are already being tested on the road in some countries such as the United States in the state of California [3]. France have also authorised the circulation of this type of vehicle since September 2022 with very strict restrictions. The shift from L2-SAE to L3-SAE implies a strong change of responsibility when driving. At L3-SAE, the vehicle is responsible for monitoring its immediate environment, using various combinations of sensors such as radar, LiDAR and cameras. The vehicle is also capable of autonomous driving when the situation allows. This means that the driver can engage in a non-driving-related task (NDRT). However, when the vehicle cannot handle a situation, it asks the driver to take over control of the vehicle, through a Take-Over Request (TOR). To do this, the driver must therefore always maintain a sufficient level of Situation Awareness (SA), i.e. perceive and understand the variables in the Driving environment at any time. In addition, increasing the level of automation in cars can lead to a decrease in trust and therefore acceptance of these vehicles. Indeed, the vehicle is the only entity monitoring the environment at L3-SAE. If it does not transmit any (or enough) information to the driver, this could reduce the driver's trust [4]. This paradigm shift poses new challenges to car manufacturers to ensure that these vehicles are fully adopted by end users while guaranteeing sufficient safety.

In particular, one of the challenges is to ensure that the driver is in good condition to take over control at any time. In the case of manual driving, factors such as fatigue (or drowsiness), distraction, mental workload, alcohol, or stress have already been shown to degrade driving performance [5], [6], [7], [8]. Drivers may already be in poor condition when they get into the vehicle, but it can also get worse while driving. At L3-SAE, engaging in an NDRT may not only increase drivers' mental load but also cause them to lose sight of the driving environment and thus reduce their SA. On the other hand, continuously monitoring the vehicle's environment during long periods of automated driving may cause drowsiness [5], [9]. To ensure that the driver is ready to take over control, it is necessary to detect these dangerous states early enough and warn the driver accordingly.

B. MOTIVATIONS AND OBJECTIVES

Given the problems that L3-SAE vehicles can engender, it is necessary to take into account the driver's state and the driving environment to induce sufficient trust and SA. The latter would contribute to optimise the takeover quality when needed. Providing optimal assistance to drivers would contribute to the safe use of conditionally automated vehicles and a reduction in the number of traffic accidents. In particular, one of the objectives is to adapt the human vehicle interaction continuously, as a TOR could occur at any time, to optimise takeover quality. Conveying Context-related information to the driver might also increase the driver's trust in the automated vehicle. Here, this information describes the status of the driving environment at a given moment. Furthermore, another objective is to evaluate the driver's state in a non-intrusive way, thus allowing acceptance and consequently trust in the automated vehicles.

C. CONTRIBUTIONS

The overall contribution of this work is the theoretical design and implementation of an innovative adaptive system to assist the driver in conditionally automated driving, and to optimize the shared control between the driver and the car. This proposed system is named AdVitam system (for Advanced Driver-Vehicle Interaction to Make future driving safer). Its goal is to adapt the interaction in the vehicle, depending on drivers' state and the driving environment's status. This supports drivers in their task of supervising the environment, but also better prepares them to take over control when needed. The driver's state is measured by physiological signals, while the state of the environment is defined by the type, Severity, and location of potential limitations. A Limitation is defined in this article as a factor that may alter the proper functioning of the automated vehicle. The taxonomy of six limitations defined by Capallera and colleagues is used as a reference for the potential hazards in a simulated driving environment [10].

The overall simplified architecture of the AdVitam system is shown in Figure 1. It is connected to a driving simulator and is designed to run continuously. The driving simulator sends information about the status of the driving environment to the AdVitam system. Physiological signals from the driver are also collected and processed by the AdVitam system and its Modules, which intelligently adapts the interaction in the car according to the received values.

The AdVitam system is composed of three modules, one of which communicates information to the other two. Each module independently constitutes an innovative contribution to scientific research:

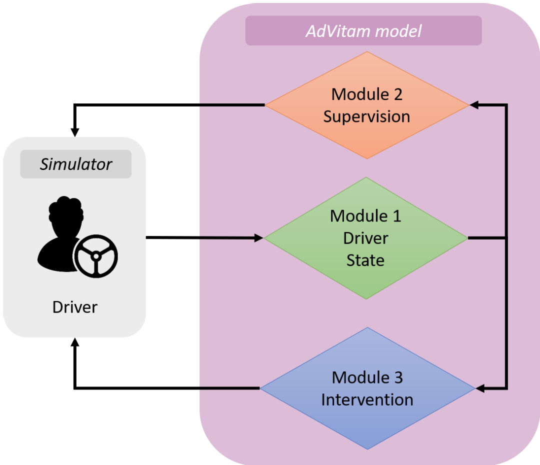


FIGURE 1. Simplified global architecture of the AdVitam system.

- ***Module 1: Driver state***. This module aims at evaluating the driver's state continuously. It takes as input three physiological signals from the driver, namely the electrocardiogram (ECG), the electrodermal activity (EDA), and the respiration (RESP). A wide range of indicators is computed from these signals to predict four risk factors for automated driving: fatigue, mental workload, affective state, and SA. These predictions are made by several machine learning algorithms. A reconstruction of the driver's global state is also proposed. The outputs of this module (i.e., values of the four risk factors and the global state) are used as inputs for the other two modules.
- ***Module 2: Supervision***. The purpose of this module is to support the drivers' supervision task by conveying context-related information through different interfaces to maintain/increase the driver's situation awareness. This module takes as input the status of the driving environment according to each category of limitation (type, severity, and location), as well as the level of mental load and the overall state of the driver. Using a rule-based model, information is sent to the driver in a tailored way to maintain/increase SA via three interfaces: ambient lights around the dashboard, vibrations through the driver's seat, and a mobile application running on a handheld device. These interfaces can be combined and operate at the same time. The choice of the modality, as well as the nature of the information transmitted to the driver, depends on the value of the variables received as input.
- ***Module 3: Intervention***. The purpose of this module is to support the drivers' intervention task when a takeover is requested, by triggering the optimal TOR modality and thus optimising takeover quality. This module takes as input the driver's physiological indicators calculated over the last 90 seconds of the collected signals (ECG, EDA, RESP) and the state of the environment (weather conditions). A machine learning model has been trained to predict the takeover quality, and to choose the best modality to request the driver to take over between three options: icon and chime, icon and vibration, or a combination of the three. The module chooses the modality that minimises the takeover quality metric (i.e., lower reaction time and steering wheel angle), based on the value of the input variables.

After a detailed presentation of the implementation of the different modules, the article also presents the implementation of the complete AdVitam system running continuously and communicating with the driving simulator. An empirical study was conducted in a static driving simulator with 35 drivers, in order to test the AdVitam system and its modules. This can be considered as a preliminary evaluation of the AdVitam system, which consists in:

- Validating the effectiveness of the AdVitam system, and specifically the ***Supervision*** and ***Intervention*** modules alone on SA, takeover quality, and trust.
- Analysing qualitatively the consistency of the predictions made by the ***Driver State*** module for both driving scenarios.

The structure of the paper follows the contributions and tasks mentioned above i.e., the design and implementation of each module as well as those of the AdVitam system. The article then details the preliminary study carried out and the results obtained. Finally, those results are discussed.

II. MODULE 1: DRIVER STATE

This section presents the literature review performed before designing the architecture of the ***Driver State*** Module. It also details the data collected and the implementation of the module performed for training the machine learning models to predict the driver's physiological state continuously.

A. RELATED WORK

In manual driving, a series of intrinsic and extrinsic factors can affect the psychological and physiological state of the driver and thus alter driving performance. Some of them might also alter the takeover performance in conditionally automated driving (L3-SAE). The aim of this section is to provide a non-exhaustive list of all the risk factors that have been shown to have an impact on driving performance in manual driving. In addition, a review of previous work that has predicted driver state is conducted, with a focus on those using physiological signals. Based on this literature review, some risk factors to be predicted for conditionally automated driving are selected in order to design the module (see Design section).

1) RISK FACTORS IMPAIRING DRIVING PERFORMANCE

There is no doubt that alcohol is one of the main causes of accidents on the roads. Significant impairment of driving skills occurs even at very low blood alcohol concentrations [11]. Alcohol mainly increases risk-taking, reaction time to danger and variability in vehicle control

[12], [13], [14]. Drugs and/or medications can also have a negative effect on driving performance, similar to alcohol [15], [16]. Other factors related to the physical condition of the driver, such as illness or poor eyesight, can also impair the driver's alertness (and therefore driving performance) [17].

Fatigue is another factor known to impair performance and cause motor vehicle accidents. It plays a role in around 10 per cent of crashes, which might be even more on highways or parkways than on other types of roads [18], [19]. Several factors impact fatigue, such as time of day, prolonged wakefulness, (repeated) sleep deprivation, or time spent on a task [20]. In manual driving, increased fatigue is often characterised by poorer lateral control of the vehicle and lower speed [21].

Distraction due to engagement in NDRT also impairs drivers' performance, primarily due to a distraction of their visual attention from the environment [22]. Engagement in another task can also affect the mental workload, which also has a clear relationship to performance and thus with the risk of having an accident [23]. The level of workload may also vary depending on driving conditions. The monotonous nature of roads can lead to performance degradation, likely due to a state of mental underload [24]. Conversely, driving in a complex environment (e.g., adverse weather and/or visibility conditions, high traffic density, etc...), can increase the driver's workload [25].

Driving conditions can also induce some driver stress, which is known to be an important factor in driver safety and individual well-being [8]. To a greater extent, drivers' affective state and emotions can have an impact on the driving behaviour [26], but also mood and personality since they are also related to affective state [17]. Other factors, such as driving experience or vehicle knowledge, can also affect driving performance. Experienced drivers generally make fewer errors and have better lateral control of the vehicle [27].

2) DRIVER's STATE PREDICTION

In the literature, many works have proposed solutions to predict the different risk factors mentioned above, using different data sources combined with machine learning techniques [28]. Fatigue and drowsiness can be predicted with high accuracy from features computed from different physiological signals or from face-related features [29], [30], [31]. Whether in driving studies or in the laboratory, brain activity has proven to be a relevant data source for predicting mental load using machine learning techniques, especially when combined with other physiological signals (ECG, EDA, RESP) or driving data [32], [33], [34], [35]. Furthermore, emotions or affective state can be predicted from physiological signals in the laboratory [36], [37]. More specifically, the prediction of driver stress from physiological signals has been the focus of numerous studies, both in simulators and in real driving conditions [38], [39]. This factor, typically occurring while driving, can be predicted with very high accuracy [39], [40].

These previous studies were conducted to predict a specific risk factor (fatigue, workload, stress, etc...) with machine learning algorithms. However, there is no model considering the combined (and continuous) prediction of several factors for the evaluation of the driver's state with machine learning techniques. The closest proposal corresponding to this problem is a model developed in the framework of a European project called Highly automated vehicles for intelligent transport (HAVEit) [41]. It included the development of a model for the evaluation of the driver's state by the combined detection of drowsiness and distraction. Their proposal was to measure long-term drowsiness and short-term distraction separately, using both direct (camera data) and indirect (driving data) measures of the driver's state [42].

B. DESIGN

1) SELECTION OF INPUT SIGNALS AND RISK FACTORS TO PREDICT

Driving data cannot be used anymore since the automated car is driving autonomously most of the time. Also, cameras might not be the most appropriate way to evaluate the drivers' state as they might not look at the windshield while engaging in an NDRT. Thus, our choice was directed towards the use of physiological signals to evaluate continuously and non-intrusively the driver state in conditionally automated vehicles. With the recent advances in wearable sensors, several signals such as EDA or photoplethysmogram (PPG) can be collected in this way through smartwatches in real-world conditions. Several measures depicting the driver's state such as Heart Rate Variability (HRV) indicators or tonic and phasic EDA indicators can be calculated from these raw signals. Thus, three physiological signals, namely ECG, EDA, and RESP, were selected as inputs for this module.

The goal is to select risk factors that impair takeover performance in conditionally automated driving, that can be assessed by physiological signals, and that are not too limiting to be experimentally manipulated. The driver might be already tired and/or stressed when getting into the car. Long periods of automated driving might also increase the driver's drowsiness and reduce workload, which might impair takeover performance. Also, engaging in an NDRT or driving in an urban area might also induce workload and stress, even in conditionally automated driving.

Besides, it may be relevant to assess the drivers' SA continuously as it might be reduced when engaging in an NDRT. It is not clear whether Situation Awareness can be assessed through physiological signals. Reference [43] suggested that it might be possible to predict SA by manipulating it experimentally and training machine learning algorithms to assess it, based on collected data. Hence, SA was considered in the module conception, to verify whether physiological signals can be considered to evaluate drivers' SA in this context.

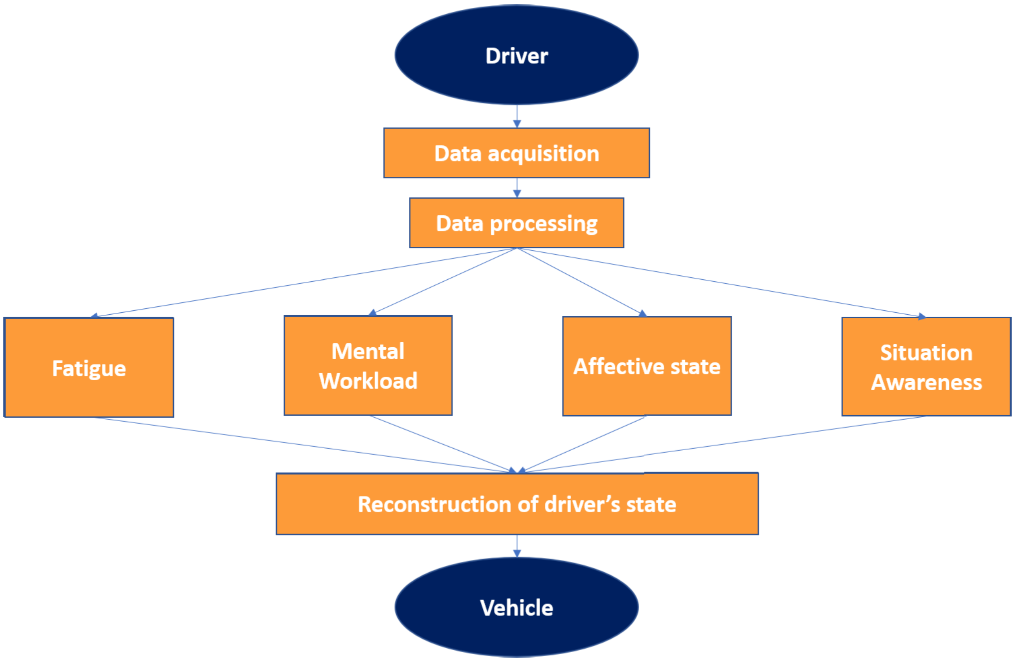


FIGURE 2. Global architecture of the ***Driver State*** module.

However, alcohol and distraction detection were not selected in the ***Driver State*** module design. The former is ethically more restrictive to collecting physiological data from drunk participants, while it is difficult to predict the latter from physiological signals alone (see the effect of task modality on classification accuracy in [35]).

Hence, four risk factors considered as critical for conditionally automated driving are selected to be predicted by the ***Driver State*** module: fatigue, mental workload, affective state and situation awareness.

2) GLOBAL ARCHITECTURE

The approach used for this module is to train several machine learning models, each one responsible for predicting one specific risk factor that can be measured through physiological signals. An overview of the ***Driver State*** module is presented in Figure 2. To provide the AdVitam system with a global indicator of the driver's physiological state, a fusion of all machine learning algorithms outputs has been designed, following the idea of the HAVEit project [41]. It could be used by other modules or by in-vehicle displays for further visualisation of the driver's state.

3) DATA COLLECTION

There are no existing datasets of drivers' physiological signals collected specifically during conditionally automated driving. In order to be as close as possible to reality, the physiological signals of drivers corresponding to the four selected risk factors were collected to train machine learning algorithms. Several experiments were designed and carried out to manipulate the selected risk factors in driving simulator experiments [35], [44]. Table 1 summarizes the three user experiences conducted.

C. IMPLEMENTATION

Figure 3 shows the implementation of the driver's state module. This pipeline was implemented in Python and is aimed at running continuously to provide the AdVitam system with an indicator of the driver's state, based on physiological data. The training of machine learning models from collected data is detailed in the first subsection. The implementation of the fusion for the *Fatigue* and *Affective State* blocks, as well as the *Global fusion* is explained in a second subsection.

1) TRAINING MACHINE LEARNING MODELS

A large amount of physiological data was collected during the different experiments conducted on the driving simulator (Table 1). Seven machine learning models were trained on the basis of physiological indicators calculated from physiological data collected in user experiences. These models predicted the four selected risk factors, all considered critical for driving: two for fatigue, one for mental workload, three for affective state, and one for situation awareness. Table 2 shows the list of all models trained and saved for further continuous driver's state prediction. All machine learning models were trained using the same pipeline implemented in Python with the scikit learn framework [45]. The pipeline is described in detail in [35] and [44].

Different segmentation levels (i.e., time windows) were tested to split the raw signals into several windows. The latter were processed separately to generate features, which allowed to provide machine learning models with more

TABLE 1. List of user experiences conducted to create the physiological dataset to train the *Driver State* module. The duration of physiological recording includes baseline, training, and driving phases.

ID	Manipulated state	Condition/Task	Participants	Duration	Ref.
#1	*Affective State* : Stress	Driving with a passenger vs. driving alone	60	25 min	[44]
	Affective State : Relaxation	Pre-driving meditation podcast vs. audiobook			
#2	*Mental Workload* : Task difficulty	N-back task : no task vs. 1-back vs. 3-back	80	70 min	[35]
	Task modality	N-back task : no task vs. visual task vs. auditory			
	Situation Awareness	Context-related information on mobile app vs. no app			
#3	*Fatigue* : Sleep-related fatigue	Sleeping time (<6 hours vs. >7 hours)	63	70 min	in press
	Affective State : Stress	Driving environment (Urban vs. rural)			
	Fatigue : Drowsiness	Beginning vs. end of drive (before TOR)			

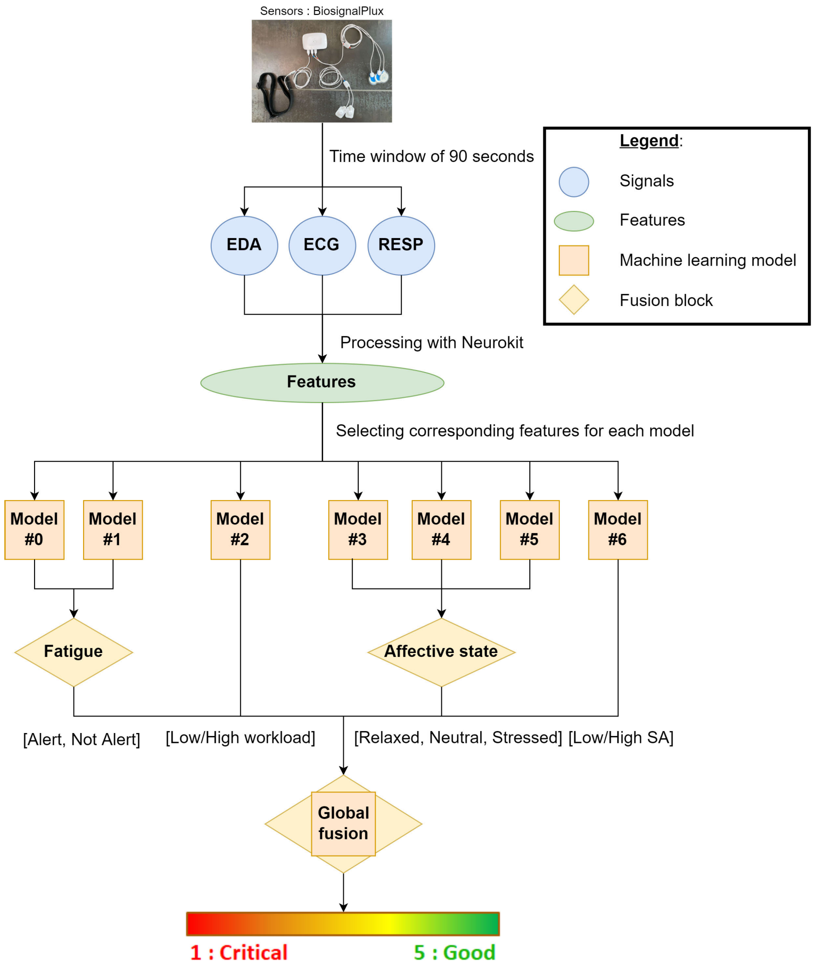


FIGURE 3. The implemented pipeline for continuous assessment of the driver's physiological state.

training samples. Raw physiological signals were processed with the Neurokit library in Python to compute a large range of physiological features [46]. For each time window, 224 features corresponding to 112 physiological indicators (10 from EDA, 74 from ECG, 21 from RESP, and 7 from RSA) were calculated. For each indicator, two features were created one corresponding to the value of the indicator while driving, and the same indicator but with baseline correction (difference between driving and baseline). Features were used as inputs of three algorithms: a random forest (RF), a k-nearest neighbor (KNN) and a neural network with one hidden layer (NN). Sensor fusion and segmentation (i.e., length of the time window used to calculate drivers' features) were tested. A repeated k-fold cross-validation approach for training and evaluating was employed for each algorithm. A weighted f1-score was used as the evaluation metric. The machine learning model that achieved the best performance over five iterations of the procedure was saved, with corresponding segmentation levels and combination of signals. To summarize the entire training process, the Table 2

TABLE 2. Summary of the selected machine learning models that gave the best performance to predict the selected risk factors. *Exp = ID of experiment (see Table 3).*

#	Manipulated driver state	Output values	Exp
#0	Drowsiness	End of rural scenario (Not alert) vs. beginning (Alert)	#3
#1	Sleep-related fatigue	>7h (Alert) vs. <6h (Not alert)	#3
#2	Mental Workload (0-20)	Low (0-10) vs. High (10-20)	#2
#3	Relaxation	Meditation podcast (Relaxed) vs. audiobook (Neutral)	#1
#4	Social stress	Passenger (Stressed) vs. alone (Neutral)	#1
#5	Stress induced by environment	Urban (Stress) vs. Rural (Neutral)	#3
#6	Situation Awareness	Mobile app with contextual info (High SA) vs. No app (Low SA)	#2

TABLE 3. Summary of the selected machine learning models that gave the best performance to predict the selected risk factors. The parameters characterising each model are reported. *Algo = Machine learning algorithm; Seg. lvl = segmentation level; Window = size of time window used to calculate features (in minute); Performance = F1-score or MAE achieved by the model (with standard deviation).*

#	Predicted risk factor	Algo	Signals	Window	Performance
#0	*Fatigue*: Drowsiness	RF	EDA	1	0.73 (0.05)
#1	*Fatigue*: Sleep deprivation	NN	EDA	1	0.99 (0.00)
#2	*Mental Workload*	KNN	EDA + ECG + RESP	1.5	3.195 (0.384)
#3	*Affective State*: Relaxation	NN	EDA	1	0.89 (0.01)
#4	*Affective State*: Presence of passenger	NN	EDA + ECG	1	0.96 (0.02)
#5	*Affective State*: Driving Environment	RF	EDA	1	0.85 (0.01)
#6	*Situation Awareness*	RF	EDA + RESP	3	0.99 (0.00)

summarizes the values predicted by each model from the experiment data (#1, #2 or #3). In addition, Table 3 shows the input signals, the time window, and the algorithm that performed best for each model.

2) RECONSTRUCTION OF DRIVER's STATE: THE FUSION

For *Mental Workload* and *Situation Awareness* blocks, only one model was trained to predict these states so no fusion was necessary. The model #2 predicted the level of driver's workload (0-10 = low workload, 10-20 = high workload) from subjective ratings obtained in the experiment #2 [35], while model #6 predicted if drivers had a high or low SA. Participants who received Context-related information on a mobile application during the drive were considered to have a high SA (see [47] for more details). For *Fatigue* and *Affective State* blocks, several machine learning models were trained to predict different components of these states.

For the *Fatigue* block, the output of model #1 (sleep deprivation) was selected when the predictions of model #0 and #1 were different. This choice was motivated by the fact that the driving scenarios to test the model in the final user experience (see section 6) are rather short (2×10 minutes) and were not designed to induce drowsiness. It was thus more relevant to detect if the driver was tired due to a lack of sleep.

For the *Affective State* block, the prediction of three models had to be considered to choose the final affective state of the driver. If the models #4 and #5 detected that the driver was stressed, the final affective state was set as *Stressed*. Otherwise, if these two models did not detect any stress and that model #3 predicted that the driver was relaxed, the final affective state was set as *Relaxed*. Otherwise, the final affective state was set as *Neutral*. This choice is debatable and the fusion of this block can easily be implemented in another way.

As shown in Figure 2, another goal was to compute a unique driver state indicator by merging the predictions made by the four blocks (*Fatigue*, *Mental Workload*, *Affective State* and *Situation Awareness*). This indicator, referred to as ''global state'', is meant to provide the AdVitam model with global information on the driver's state used to properly adapt the in-vehicle interaction for maintaining awareness as well as optimize take-over-requests. This indicator is defined as a continuous value between 1 and 5. Lower values mean worse driver state conditions with respect to fatigue, mental workload, affective state and situation awareness criteria. It can be defined as a decision-based fusion (or late fusion) approach as it makes a prediction from values returned by the upper-level blocks [48]. A machine learning model was used to remain consistent with the rest of the architecture module. A questionnaire was created and sent to the authors' academic institutions to build the dataset and create an initial ground truth. Respondents were asked to rate the driver's state and ability to drive a conditionally automated vehicle (L3-SAE) on a 5-point Likert scale (1=very bad, 5=very good), depending on the driver's state regarding each component: fatigue (low/high), mental workload (low/high), affective state (relaxed/neutral/stressed), and situation awareness (low/high). 34 respondents rated the 24 possible combinations ($2 \times 2 \times 3 \times 2$) resulting in a dataset of $34 \times 24 = 816$ samples. A linear regressor was trained to predict a continuous value between 1 and 5, based on ground truth collected from the questionnaire. The scikit-learn framework [45] was used to implement it, and a standard train/test split was done for training the model.

III. MODULE 2: SUPERVISION

As a context reminder, with L3-SAE vehicles, drivers are necessary but do not have to always monitor their environment.

They must be ready to take over the control of their vehicle at all times with notice. Because drivers will not be fully engaged in driving, vehicle disengagement will generate a higher cognitive load. This section explores the potential of using full-body and multi-sensory experience by considering the whole car interior and peripheral interaction in order to support the drivers' supervision task by conveying context-related information through different interfaces.

A. RELATED WORK

1) SITUATION AWARENESS

Situation(al) Awareness (Situation Awareness) is the perception of environmental elements and events with respect to time or space, the comprehension of their meaning, and the projection of their future status [49]. In the context of driving, this means that drivers perceive information about their environment and the functioning of their vehicle that they can understand and interpret to mentally project this element into the future. For example, they can perceive the current speed of their vehicle and the traffic around them. They can then understand the difference between their speed and the speed of the vehicle in front of them to project the position of their vehicle in the future and finally make the decision to brake if the distance is too short [50].

SA may be impacted positively by the driver's good training, experience and abilities. However, SA may also be impacted negatively by different factors such as a high cognitive workload, fatigue, and stress by carrying out an NDRT. This can lead to drivers getting out of the supervisory loop, having less information about the vehicle's environment, and could consequently affect their ability to properly regain control of the vehicle when necessary. This phenomenon, called ''out-of-the-loop'', has been introduced and identified in the literature for a few years now [51]. Moreover, the lack of active involvement and the highly automated driving could become problematic by inducing drowsiness and underload if there is no associated secondary task [52] and thus impairs SA.

2) HUMAN-VEHICLE INTERACTION MAINTAINING SA

Bakker's research on the ''interaction-attention continuum'' [53] shows the effectiveness of peripheral interactions to inform users of their environment even while they are engaged in another task. This type of interaction could help drivers in their (passive) supervision task. Many works demonstrate the effectiveness of using human-vehicle interaction to maintain driver SA using vibration in the driver's seat [54], [55], ambient lights in the periphery of view [56], [57], text and icons on the windshield [58], [59] or even audio-visual interaction [60], [61]. To date, there are few concepts that attempt to combine several modalities as well as the whole vehicle. For example, [62] combines vibrations in the seat and peripheral lights. In this article, only the environment and the status of the vehicle are taken into account. However, we have seen that the drivers' state can also impact their SA. It is therefore important to take it into account. Thus, we propose interfaces that adapt in terms of modality and/or location according to the driver's state and to the driving and interaction context (environment and NDRT) in order to efficiently maintain the driver's situation awareness. Indeed, conveying Context-related information seems a promising strategy according to the article proposed by [47].

B. DESIGN

The Human-Vehicle Interaction (HVI) for supervision combines haptic interaction (vibrations in the seat) and two visual interactions (ambient lights and mobile application) described below and illustrated in Figure 4. The design and implementation of the different interfaces are detailed in the article [63]. The study shows a positive impact of the use of these interfaces on SA.

- **Haptic Seat**: Transmits the presence of an obstacle around the vehicle (continuous vibration) and the state of right and left lane markings (discontinuous vibration). The intensity of the vibration reflects the Severity of the Limitation. The location of the limiting factor in the environment corresponds to a particular zone of vibration in the seat. For example, the presence of an obstacle at the front left of the vehicle will be indicated by a vibration at the front left of the bottom seat (i.e., under the thighs). The danger of the obstacle will be reflected by a continuous vibration more or less intense.
- **Ambient lights**: Reflects the general severity of the environment. They are displayed in the driver's peripheral field of view. A green colour means there is no limitation in the environment. Yellow corresponds to a low severity while orange corresponds to a more important severity but which still does not require a takeover.
- **Mobile application**: An icon appears on the screen if the vehicle encounters a limitation. The drawing represents the limitation and the colour of its severity. It is possible to click on the icon to have more information. Any change in severity is notified by the appearance of the icon followed by a double horizontal back and forth from right to left.

The choice of colours, intensities and combination (or not) of interfaces is driven by a rule-based model. This system takes into account the driver's state given by the ***Driver State*** Module described before as well as the limitations in the vehicle environment based on the taxonomy from [10] (Environment, External Human Factors, Road, Lane, Obstacles and Vehicle Alteration). The implementation of this rule-based model is defined in the following section with an emphasis on the multimodal aspect of the interaction.

C. IMPLEMENTATION

The rule-based model takes into account the following input parameters (see Table 4): the type, the severity and the location of the limitation, the engagement ($10 < mentalworkload \leq 20$) or not ($mentalworkload \leq 10$) in a

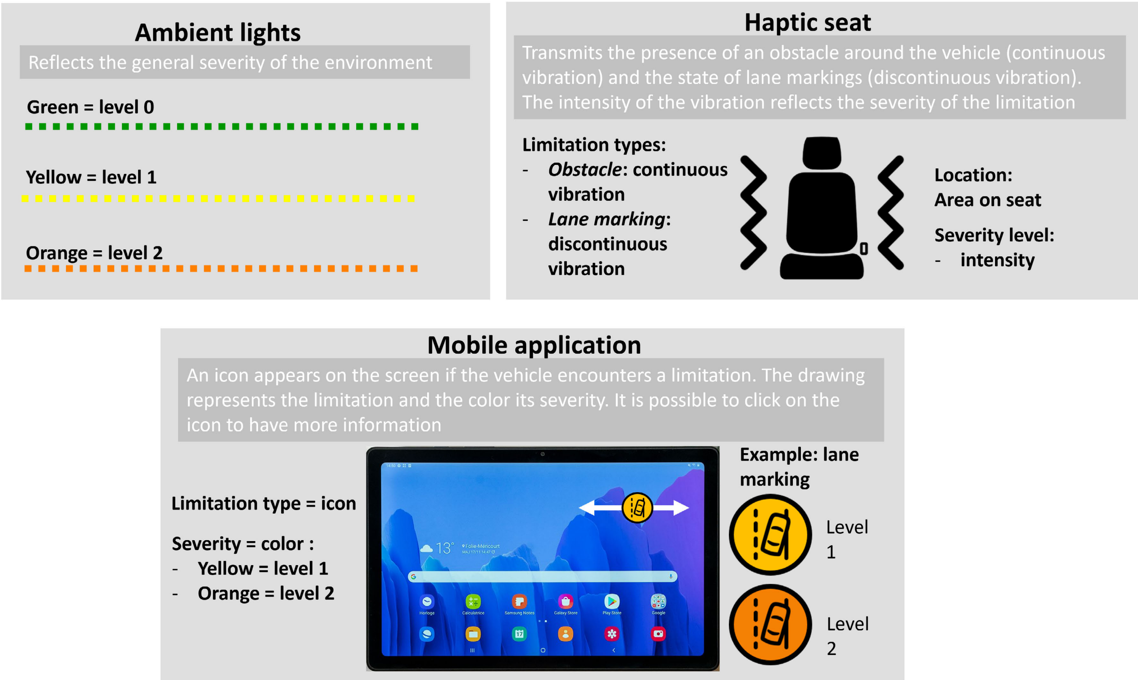


FIGURE 4. Interfaces and modalities managed by the ***Supervision*** module.

TABLE 4. Summary of the inputs on the rule-based model.

	Input
Limitation	**Type**: Environment, External Human Factors, Road, Lane, Obstacles and Vehicle Alteration
	Severity: 0, 1 or 2
	Location: both, left, right
NDRT	Yes (1), No (0)
Driver's state	Good (1), Bad (0)

NDRT and the state of the driver (good for $globalstate > 3$ or bad for $globalstate \leq 3$). The first inputs are sent by the simulator while the engagement in the task and the driver's state are received from the ***Driver State*** module. The output modalities (see Figure 4) used to sustain the driver's SA are the haptic seat (vibration location, pattern and intensity), the ambient lights (colours) and the mobile application (icons drawing and colour). Figure 5 summarizes the set of rules implemented for this module.

The implementation of this module uses the durable-rules framework in Python. Thus, taking the example of erased lines (''Lane'' category) of severity 2, in case the driver does not perform an NDRT and his/her state is considered as bad, the following rule would be selected by the module:

if $(Lane\_severity = 2)$ & $(driverstate = 0)$ & $(NDRT = 0)$) **then**
set_modality(*lane*, *modality*, *characteristics*)
end if

with *modality* = [light & seat] and *characteristics* = [orange ambient light, high intensity vibration]

In this case, the set_modality() method would return the category (Lane) and the characteristics of the concerned interfaces (orange ambient light and high-intensity vibration in the seat). Let's take another example and assume that a rock (category ''Obstacle'') is detected by the vehicle, rated with a severity of 1. If the driver performs an NDRT, but its overall state is still considered good, the following rule would be triggered:

if $((obstacle\_severity = 2)$ & $(driverstate = 1)$ & $(NDRT = 1))$ **then**
set_modality(*obstacle*, *modality*, *characteristics*)
end if

with *modality* = [light & mobile application] and *characteristics* = [yellow ambient light, yellow obstacle icon]

To summarise, this module considers the limitations of the environment provided by the simulator as well as the global state of drivers and their cognitive state (i.e., performing an NDRT or not). These are defined by the ***Driver State*** module. A rule-based model chooses a combination of interfaces to use within the vehicle. The characteristics also depend on the input parameters. The different combinations chosen are intended to maintain the driver's SA but also their trust in the vehicle during the autonomous driving phases. The choice of

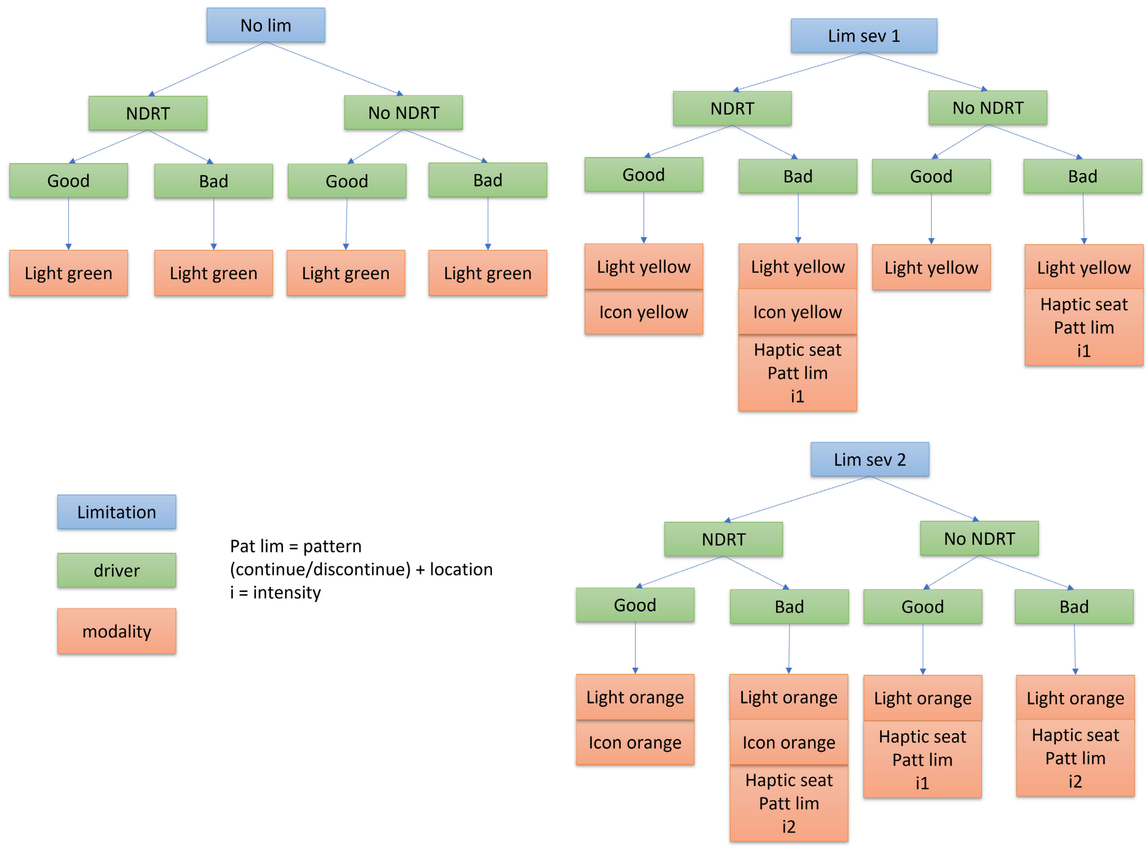


FIGURE 5. General logic of the rule-based model.

using the mobile application is only made when the driver is performing a secondary task because it is the same device. The lights reflect the overall severity as this seems to be effective since the literature mentioned above. Finally, the choice of using the haptic seat and different intensities is established following the study from [63]. The proposal of this model and the validation of these choices are evaluated in the rest of the paper

IV. MODULE 3: INTERVENTION

This section explores the potential of using machine learning to predict takeover quality, and use this information to support the driver by providing the most suitable takeover request modalities for any given takeover situation. The related work section highlights the literature that investigated the impact of takeover request modalities on takeover quality, as well as research that proposed approaches to predict takeover quality. This allows us to transition to our proposed agent design and implementation.

A. RELATED WORK

1) THE IMPACT OF TAKEOVER REQUEST MODALITIES ON TAKEOVER QUALITY

Supporting the driver for takeover situations is usually done prior to the takeover by raising the driver's awareness or forcing the driver to monitor the environment periodically. Several works were done to see the impact of specific modalities on takeover performance, with a few conclusions but no consensus on a perfect set of modalities.

In particular, [64] showed that multimodal signals were perceived as more urgent by the drivers and allowed drivers to take over faster on straight roads, but their reaction time (RT) decreased on curved roads. In the case of unimodal takeover requests, visual was repeatedly shown as being the less efficient unimodal takeover request, with higher reaction time than auditory or haptic modality [65], but those performances could be significantly increased by adding the auditory modality, according to [66]. Reference [67] showed that a static pattern for the haptic modality led to a slightly shorter reaction time as opposed to a dynamic pattern in the haptic driver seat. Reference [68] showed that having the haptic modality in the driver seat also help for takeover when the expected behaviour from the driver was to change lanes. However, [69] did not see any improvement in reaction time using an experimental haptic steering wheel. Reference [70] concluded by showing that the haptic modalities for a takeover request should be located on the seat rather than the steering wheel.

In the case of experimental modalities, [71] proposed the concept of Useful Field of View (UFOV), which extends the field of view of the driver with augmented reality and provided icons during takeovers. However, if the results showed that UFOV was beneficial for experienced drivers, with a significantly lower reaction time, the more inexperienced drivers displayed a higher average reaction time.

2) MACHINE LEARNING STUDIES ON TAKEOVER QUALITY

Most of the work on conditionally automated vehicles and artificial intelligence (AI) focus on driving intelligence, using sensors and various captor. However, some research was done specifically on using AI to improve the takeover quality, which is highlighted hereafter.

Reference [72] proposed a predicting model of the takeover scenario risk level, which was a label deducted from the takeover performance of the driver. In a sense, it is indirectly predicting the takeover performance of the driver. There were in this case three possible risk levels, which an RF algorithm was able to predict with an accuracy of 98.8%. Reference [73] proposed a regression formula to estimate the reaction time of the driver after a takeover request. This highlighted the potential features to consider as well as showing that regression of a takeover quality metric is indeed possible. The time between the takeover request and the estimated time before a collision, the lane position of the car, the traffic density and the number of times the drivers encountered a similar situation were all features of the regression. The age of the driver was shown to have a minimal impact on the reaction time. Reference [74] proposed a model which aims to predict the takeover quality of the driver using the external environment and the driver state. In their study, the takeover quality was a binary label, which was either "good" or "bad". This binary estimation was done by the researchers based on the overall quality of the takeover. The best performing model was a Random Forest with an f1-score of 64%. Reference [75] proposed a regression formula which was already discussed and refined in [73], but proposed a regression model for the lateral acceleration as well, a metric closely related to Maximum Steering Wheel Angle (MaxSWA), one of the metrics used in this piece of research. The variables are the same as the ones presented in the previous regression, showing that the same features can be used to predict both metrics.

B. DESIGN

Focusing on the highlighted points from the previous section, a smart Module using AI is proposed to support the driver in a takeover situation. This module aims to improve the takeover quality by providing the most suitable takeover modalities for the takeover request. It essentially functions as a human-machine interface able to design its interactions with the driver on the fly based on the situation.

In order to do so, the module needs to consider pertinent inputs from which it can extract relevant features, allowing it to model the takeover quality. Once this step is validated, it can infer variations in the takeover quality caused by the modifications of the inputs. This means the module design needs to resolve the following points:

1) Identify the factors influencing takeover quality.
2) Choose the factors that the module can act upon to directly influence this takeover quality.
3) Model the takeover quality according to the literature metrics.
4) Recommend modifications on the chosen factors to increase the takeover quality.

1) RELEVANT FACTORS

The literature and modules presented in the previous sections highlighted four potential inputs for modelling the takeover quality:

- The driver's physiological state.
- The driver situation awareness.
- The external environment.
- The takeover request modalities.

The driver's physiological state is a crucial source of information directly derived from the module 1 ***Driver State***. Driving performance is influenced by the mental state of the driver, and knowing this state should allow modelling more precisely the takeover quality. Both the driver's physiological features and state are potentially good sources of information, and should both be taken into account if possible. Driver Situation Awareness is a construct that is difficult to measure continuously with physiological signals but was shown to influence takeover quality [76]. An approach is proposed in Module 1 to investigate whether SA can be inferred with high accuracy. However, the system needs to be validated before its prediction can be used for this module. As such, it will unfortunately not be considered as an input for this module. The external environment was shown [77]to impair takeover quality especially when it consisted of adverse weather, and should be monitored. The takeover request modalities influence heavily the takeover quality and are the only factor that is directly decided by the vehicle, making it easy to monitor and change.

By analysing the four factors presented above, the takeover modalities appear to be the only factor that can be easily controlled to influence the takeover quality. The modalities chosen for this module are visual, auditory and haptic, which were the modalities highlighted by the previous section 4.1.1 "The impact of takeover request modalities on takeover quality".

2) MODELING THE TAKEOVER QUALITY

As proposed in the previous section, the module aims at modelling the takeover quality. In order to achieve this goal, a machine learning model must be developed where the inputs are used to predict the takeover quality, before making a recommendation on the most suitable takeover request modalities using this model. Figure 6 shows the architecture of the module. This architecture is an extension of the one proposed in [78], which only proposed a model of takeover

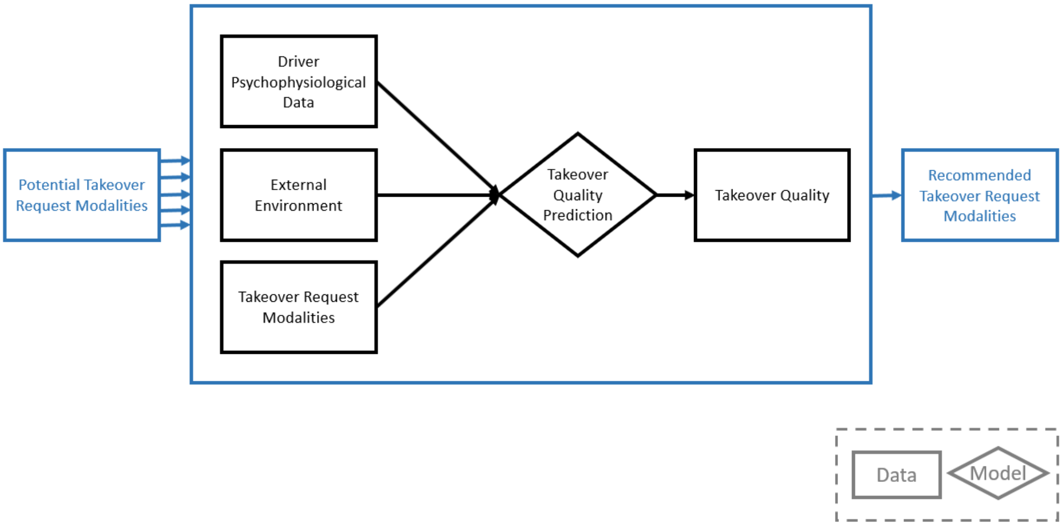


FIGURE 6. Architecture of the ***Intervention*** module, with the extension from previous architecture in blue.

quality. Here, the potential takeover modalities are all tested to allow for the selection of the best-performing one, which is then recommended to the system.

The following section focuses on the implementation of this design.

C. IMPLEMENTATION

In this section the implementation of the module ***Intervention*** is discussed. There are two major sections: the machine learning methodology, including the models training and evaluation, and the real-time implementation of the best performing models, which is the version we propose in this piece of research.

1) TRAINING AND EVALUATING MODELS FOR TAKEOVER QUALITY PREDICTION

The machine learning model created by [79] was used for the takeover quality prediction. In a user experience, the physiological data of 15 participants were collected continuously. In particular, the physiological indicators corresponding to the state of the drivers 90 seconds before a takeover situation were calculated to create the dataset. A total of 80 takeover situations could be tested. Drivers' takeover quality was deducted using the normalized aggregation of the driver RT and MaxSWA attained during the takeover process. Several machine learning models were trained to predict this takeover quality using the driver's physiological features from the 90 seconds prior to the takeover request, as well as the external environment (represented by the weather condition) and the takeover request modalities.

We can see in Table 5 all the trained machine learning models and the baseline, which was the constant prediction of the average of the takeover quality across all takeovers. Support Vector Regressor appeared as the best performing

TABLE 5. Summary of Mean Squared Error (MSE) and Mean Absolute Error (MAE) scores of models on the takeover quality prediction. Best model in bold.

Algorithm	MSE	MAE
Baseline	0.6000	0.2073
KNN	0.0691	0.2013
Support Vector Regressor	**0.0538**	**0.1614**
RF	0.0883	0.2421
Multi Layers Perceptron	0.0779	0.2478

model, and was then selected for the next step: modalities recommendation.

The recommendation quality was then evaluated using the methodology described in [80]. There were two evaluation criteria: the modality impact and the modality diversity. Modality impact ensures that switching modalities does indeed influence the takeover quality prediction, as it should, according to the literature. Modality diversity ensures that no modalities are systematically better than others, which should not happen, according to the literature. Modality impact is calculated as the maximum difference between each modality prediction of the takeover quality, while modality diversity is a statistical check of the distribution of the recommended modalities. These evaluation criteria were done only on the test set of the dataset, in order to evaluate their impact on unseen data and avoid bias as much as possible.

The modality impact of the Support Vector Regressor was 4.95% on average, with a standard deviation of 2.7%. This means that on average the difference between the best performing modalities and the lowest performing modalities was a little less than 5%. Regarding modality diversity, the most recommended modalities were the auditory-visual combination, with 62.5% recommendation (10 out of 16). The haptic-auditory-visual combination was second with 25% of the recommendation (4 out of 16) and the haptic-visual

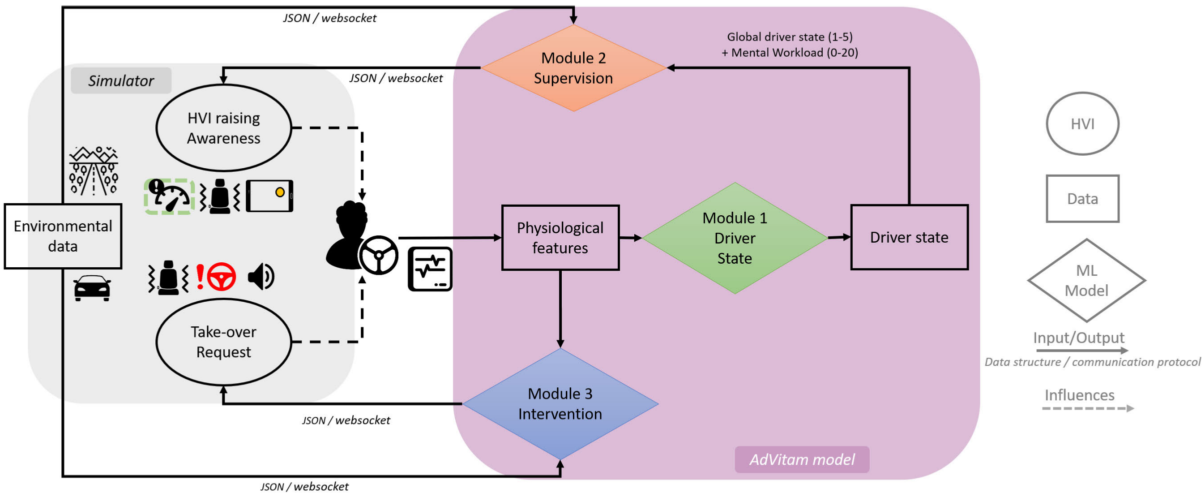


FIGURE 7. Global architecture of the AdVitam system.

combination represented 12.5% of the recommendations (2 out of 16).

2) CONTINUOUS PREDICTION OF OPTIMAL TOR MODALITY

To be able to function properly in the AdVitam system system, this module needed to be adapted to operate continuously. This meant receiving the features, predicting the takeover quality for every possible set of modalities and recommending the modalities associated with the best takeover quality prediction. This loop was scheduled to run every second, to always have a recommendation ready in case of a need for a takeover request. There were several steps needed to transform the model presented in the previous section to work in this new context. This process is detailed hereafter.

Despite working well in the previous conditions, the model was not able to transfer flawlessly to this new context. The modality impact criteria from the previous evaluation dropped to 0 in the first test of the whole setting, which forced the original design to be modified to be more robust. A secondary model, named the backup model, was added to the prediction pipeline in order to make a secondary prediction if the first one fails to do so. Failure was defined as having a modality impact of 0%, meaning each modality returns the same takeover quality prediction and no recommendation can be made. The second model was a KNN model which was trained on the same data presented as the first model, but with the first model prediction as the label. This ensures consistency in both models' prediction, with the second model being able to predict directly the recommended modality instead of the takeover quality. The combination of both models proved to be able to make a prediction in most cases and was deemed robust enough to operate continuously with the other modules, which is presented in the following section.

V. THE AdVitam SYSTEM

This section presents the theoretical architecture of the AdVitam system system and its implementation. It has been implemented in order to be able to operate continuously and to be tested during a driving simulator study with drivers.

A. THEORETICAL ARCHITECTURE

The global architecture of AdVitam is presented in Figure 7. It shows the interaction between the three Modules, but also between the driving simulator and the modules. The physiological signals of the last 90 seconds are first collected and then processed by the Python module. The indicators corresponding to this time window are then transmitted to modules 1 and 3. From these indicators, module 1 (***Driver State***) is executed to predict the driver state (four dangerous states and global state, see Figure 3). Module 2 (***Supervision***) retrieves the prediction of the mental workload (high or low) and the global state of the driver from module 1, as well as the environmental data from the simulator. Based on this, it triggers the appropriate in-vehicle interfaces to maintain driver awareness. The Figure 5 summarizes the modality of the information transmitted to the driver according to the input values (see Table 4). Module 3 (***Intervention***) takes as input the same physiological indicators as module 1, with additional information on weather conditions (good or bad weather). Based on this, it predicts the optimal takeover modality using machine learning techniques (see Figure 6). Depending on the highest Severity in the environment, the output of module 2 or 3 is selected for in-vehicle interaction (see Figure 8 in the next section).

B. IMPLEMENTATION

To implement the theoretical operation of the AdVitam system explained above, several processes were implemented to run in parallel:

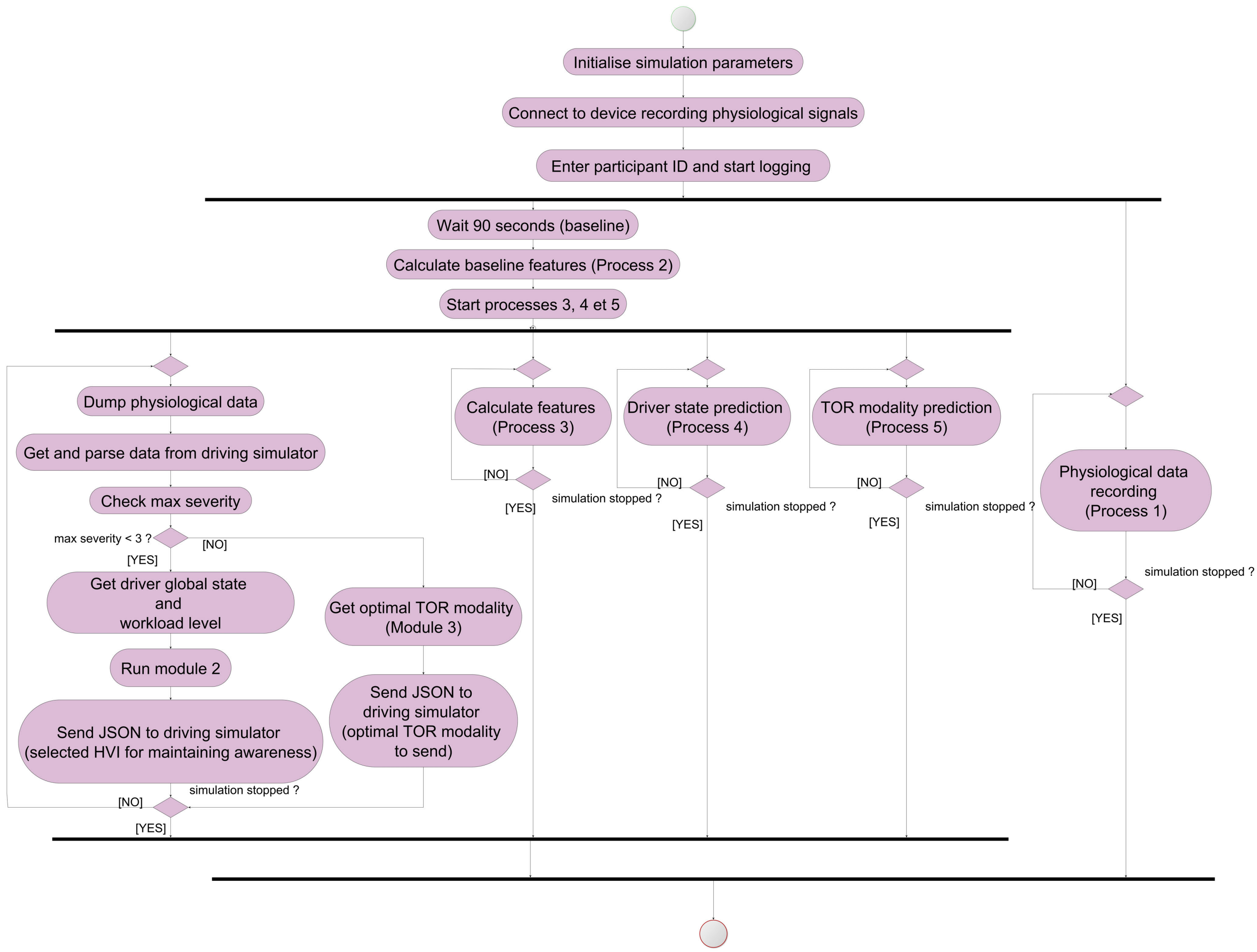


FIGURE 8. Flowchart (activity diagram) of the Python module operation.

- Main process: Starts the main loop, the others process and receives/sends data from/to the driving simulator
- Process 1: Raw physiological data acquisition (with calibration)
- Process 2: Calculation of physiological features after baseline (90 seconds)
- Process 3: Calculation of physiological features during automated driving (last 90 seconds)
- Process 4: Driver state prediction (module 1, scheduled to run every second)
- Process 5: Take-Over Request modality recommendation (module 3, scheduled to run every second)

Figure 8 shows the functioning of the Python module. The main process starts by initialising the simulation parameters (participant code, sampling rate, time window) and connecting to the driving simulation via WebSocket. Process 1 starts the acquisition of raw physiological signals in parallel. The raw values of the EDA, ECG and RESP signals from the last 90 seconds are saved at each iteration of the main loop. Once the first 90 seconds have elapsed, Process 2 processes the raw data. The driver's physiological indicators during the baseline are calculated with the Neurokit library [46] and saved to a CSV file for later use.

Next, processes 3, 4, and 5 are started to run in parallel to the main loop. Process 3 calculates the driver's physiological indicators while driving (*Dr* features). Those calculated after process 2 are also used by process 3 to generate additional features corrected with the baseline (*Dr-Bl* features). For this, the difference between the two values is made, as it is the case for the driver state module. All calculated physiological features are saved in a CSV file. They are used by processes 4 and 5 to respectively predict the driver state and the TOR modality. These two processes are scheduled to be executed every second.

Once all processes are started, the main loop iterates until the driving simulation is stopped. The last frame sent by the driving simulator, the global state of the driver as well as its workload level (process 4), and the TOR modality (process 5) are retrieved by the main loop. For the latter two, default values are assigned if the prediction could not be executed:

10 for the workload level, 3 for the global state of the driver, and the audio-visual modality for the TOR.

Then, the data received from the driving simulator is parsed. The data describing the state of the Driving environment is modelled by a frame in JSON format, containing the micro-category, the severity and the location of the five macro-categories of potential danger. The frame sent to the driving simulator is structured in a JSON format.

Then the main loop calculates the maximum severity among the five categories. Based on this, the outputs of modules 2 and 3 are fed back to the driving simulator. If the maximum severity in the environment is lower than 3 (see Figure 8), the output of module 2 (supervision) is retrieved and the HVIs for maintaining/increasing situational awareness (Situation Awareness) are used. The data frame sent to the driving simulator is also structured in JSON.

If the highest severity is 3 (see Figure 8), a TOR must be triggered to request the driver to regain control. In this case, the last output of module 3 (i.e. the optimal TOR modality) is retrieved, sent to the simulator and immediately transmitted to the driver. The frame sent to the driving simulator is also structured in JSON.

Throughout the simulation, logs of predictions made by the driver state module (each model, each block, and the global driver state), the intervention module (takeover modality), and the supervision module (rule selected by the model) are saved. All the physiological indicators of all participants calculated after each iteration of process 3 are also saved.

VI. EMPIRICAL STUDY

This section describes the research hypotheses of the empirical study as well as the experimental design and the procedure put in place to evaluate them. It also describes the materials, instruments, measurements, and statistical analysis used during and after the experiment.

A. GOALS AND HYPOTHESES

The purpose of this experiment is to be able to test the complete AdVitam system in a simulated driving situation with participants. In particular, the goal is to evaluate the effectiveness of each of the Modules on different safety-related measures, namely SA and takeover quality, but also measures of trust, mental workload, fatigue, and user experience. The hypotheses are as follows:

- ***(H1)***: Module 1 should return a prediction consistent with the driver's experimental manipulation. Thus, we should observe that mental workload increases during the NDRT phases (closer to 1) [7], that affective state is closer to stress during the urban scenario (closer to 2 on the 0-2 scale) [38], and an increase in SA for drivers receiving Context-related information on the environment [47] (***Supervision*** condition, see the experimental design below). In addition, the overall state of the driver should be worse during the NDRT phases and we should not see a major change in fatigue due to the short duration of the experiment.
- ***(H2)***: Receiving contextual information about the car's environment through the multimodal interfaces (***Supervision***) should increase (or at least maintain) SA and tend to improve takeover quality [47], [76]. It should also increase user trust and experience in the car as it increases the vehicle's transparency [62].
- ***(H3)***: The intelligent adaptation of the TOR modality (***Intervention***) should improve the takeover quality [66], [67]. This should also increase the user's trust and experience in the car as it increases the transparency and personalisation of the vehicle.
- ***(H4)***: Driving with the full AdVitam system (***Supervision*** $\times$ ***Intervention***) should not increase participants' mental load and fatigue (i.e., sleepiness). It should also increase driver trust and user experience, consistently with hypotheses 2 and 3.

B. PARTICIPANTS AND DESIGN

35 participants (11 females) were recruited for the experimental evaluation. The age of the participants ranged from 19 to 55 years old ($M = 26.1$ years old; $SD = 65.98$ years old). On average, they reported driving 6305.43 km per year ($SD = 88421.53$ km) and have held a driving license for about 7.4 years ($SD = 6.54$ years). The only criterion for participating in this experiment was to be in possession of a valid driver's license. Students received course credit for their participation. All participants were entered into a draw to win one of five 20 CHF vouchers for a store in the city. All the research and measurements followed the tenets of the Helsinki agreement and consent form was obtained from all participants.

The design of the experiment included two between-subjects factors (***Supervision*** and ***Intervention***) as well as one within-subjects factor (rural vs. urban scenario). The first between-subjects factor was the presentation and use of the supervision module. Drivers received information about the environment from in-vehicle interfaces according to the situation and their state. The second one was the presentation and use of the intervention module. Drivers received different takeover modalities (audio-visual, haptic-visual, and audio-haptic-visual) with regard to the environment of the car and their own state.

Throughout some parts of the driving session, participants had to perform a cognitive NDRT on a handheld tablet (visual 2-back task). In addition, the participants had to react to the presence of five Limitations and two takeover situations divided into a rural and an urban scenario. The order of apparition and the location of limitations and takeover situations in the scenarii was the same for all participants. Figure 10 describes the order of occurrence of the limitations, the times during which drivers perform an NDRT, and the situations leading to a TOR. Each takeover was requested due to an issue with automation. These situations depicted three of the six categories from the taxonomy [10], which were:

FIGURE 9. A participant in the driving simulator during the experiment.

Rural scenario:

- A slightly damaged lane marking (*Category Lane*)
- A more slightly damaged lane marking (*Category Lane*)
- A rock on the other lane (*Category Obstacle*)
- Rocks blocking the way (*Category Obstacle*)

Urban scenario:

- A slightly damaged lane marking (*Category Lane*)
- A dog standing on the right side of the road and then crossing over (*Category External Human Factor*)
- A pedestrian walking very close to the road (*Category External Human Factor*)

C. MATERIAL, MEASURES, AND INSTRUMENTS

The experiment was conducted on a driving simulator specifically built for the project. Rural and urban environments were both used for the experiment. All participants were given a Samsung Galaxy Tab A (10'') to perform the NDRT. An Android mobile application was developed to administer the visual 2-back task. Figure 9 shows a participant sitting in the driving simulator, performing the non-driving related task during the experiment.

Physiological signals from the drivers were recorded using the Biosignals PLUX kit at a sampling rate of 1000 Hz. This value was chosen to have the same data frequency as for the data collected in the previous experiments. The data were sent in real-time via Bluetooth to a laptop on which the AdVitam system was running. Pre-gelled disposable Ag/AgCl electrodes (EL507 and EL503, Biopac) were connected to the kit's wire sets. ECG electrodes were placed on the left side of the drivers' stomachs, following the Lead-I configuration. EDA electrodes were attached to the index and middle fingers of the left hand. The breathing belt was attached around the abdomen at the stomach region. Before each scenario, drivers' arousal, valence, and sleepiness were evaluated through questionnaires presented on a tablet. After each scenario, they were also assessed in addition to the mental workload, SA, and trust towards the vehicle during the drive.

Valence and arousal were assessed on a 1-5 scale using the animated version of the Self-Assessment Manikin (AniSAM) [81]. Sleepiness was self-rated by participants on the Karolinska Sleepiness Scale (KSS) [82]. The mental workload was evaluated on a 0-20 scale using the Mental demand item of the NASA-TLX questionnaire [83].

Different techniques exist to measure awareness, such as objective, subjective, performance-based or process index-based methods. In our study, the drivers' Situation Awareness was assessed using the Situation Awareness Rating Technique (SART) [84] and an open question asking for the cause of the TOR. Post-trial questionnaires were used because we preferred not to freeze the simulation while collecting physiological data. Each SART item was rated on a 7-point scale, from 1 (Low) and 7 (High). Trust was assessed using the Situational Trust Scale for Automated Driving (STS-AD) [85]. A trust score was calculated for each scenario, according to the authors' instructions.

To measure objectively drivers' SA, measures of task performance and identification rate of limitations were calculated. The task performance on the visual 2-back task was calculated according to the formula:

$$TaskScore = \frac{correct\ answers - wrong\ answers - missed\ targets}{correct\ answers + missed\ targets} \quad (1)$$

To assess whether participants saw and understood the limitations that occurred during the experiment, a think-aloud method was used. They had to orally report what type of limitation they saw in the vehicle's environment, with its Severity and location. Their answers were transcripted during the experiment. During data preparation, participants' answers were coded separately by two authors. Each limitation was scored separately, with the type, severity and location. One point was accorded if the participant answered correctly. All limitations occurring in the experiment (see Figure 10) were scored, except for the two takeover situations and the pre-alert in the rural scenario (Obstacle, severity 2, front). Cases, where the two authors disagreed, were discussed afterwards to make the final decision. For each scenario, an average score was computed for limitations' type, severity and location separately.

The quality of the takeover can be measured through different metrics that can be categorized into five categories [86]: time/distance margin, speed, offset in lane, steering and brake behaviour. In this study, takeover quality was evaluated from the driving data. The metrics used were the drivers' RT and MaxSWA. The number of collisions was also reported (failed Take-Over Request). We computed these three metrics because the experiment combines situations where the driver can either avoid an obstacle or brake to complete a stop.

At the end of the experiment, participants filled in the User Experience Questionnaire Short version (UEQ-S) [87] in order to assess their user experience in the driving simulator.

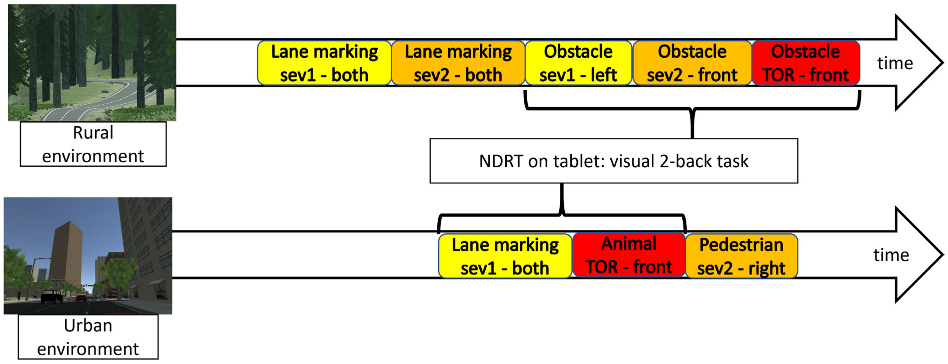


FIGURE 10. The experimental procedure.

D. PROCEDURE

After initial instructions about the experiment, the participants filled in a questionnaire on a tablet containing socio-demographic questions, as well as questions about their driving habits and a consent form. Each participant had to familiarize themselves with the limitations of automated vehicles. Participants were presented the taxonomy with categories and an illustrated situation involving all categories. They also received instructions about modalities according to the module they encountered (***Supervision*** and/or ***Intervention***). From that point, the experiment was divided into three distinct periods:

- **Training session**: participants received instructions on the operation of the autonomous pilot (activation, deactivation, TOR) and on the interactions of the simulator according to the test conditions. They then practised driving the simulator for a few minutes to become familiar with the equipment. The device for recording physiological signals with electrodes was installed at the end of this phase.
- **Rural environment**: this session started with a two minutes baseline where participants only monitored the environment. Then, they experimented with the rural driving session. Their task was to mention verbally the limitations they may encounter on the route. They had to also perform an NDRT (visual 2-back task) on some parts of the route. In the end, they filled out SART, AniSAM and KSS questionnaires.
- **Urban environment**: During this period, they experimented with the urban driving session. The instructions were the same as in the rural scenario.

In the case of a TOR, all participants were instructed to react accordingly and drive the car manually until the critical situation was over. Once they had estimated that the situation was safe again, they could activate the autopilot again. If they forgot to reactivate, the researcher reminded them.

At the end of the session, participants were asked to stop the car and leave the simulator to fill in the last part of the questionnaire. Finally, participants were thanked and discharged.

E. STATISTICAL ANALYSIS

A repeated-measures analysis of variance (ANOVA) was run on measures of workload, SA, trust, task performance, identification rate of limitations, and takeover quality, with the ***Supervision*** and ***Intervention*** modules (with vs. without) as between-subjects factors, and the Driving environment (rural vs. urban) as a within-subject factor.

For measures of arousal, valence and sleepiness, the driving environment was replaced by the measurement time as the within-subjects factor, but with three levels (baseline vs. rural vs. urban). To investigate the effect of the driving environment, ***Supervision*** and ***Intervention*** modules on the number of crashes, a Fisher's exact test was done to test the hypothesis that the two column percentages in a 2×2 table are equal.

For user experience measures, only an ANOVA with ***Supervision*** and ***Intervention*** modules was run, since it was only measured once at the end of the experiment.

VII. RESULTS

This section describes the overall results obtained during the study. This part allows us to evaluate the impact of the AdVitam system system and of the different Modules on the drivers' state, SA, trust and user experience. In the following paragraphs, M stands for Mean, SD for Standard Deviation and SE for Standard Error.

A. AROUSAL, VALENCE, SLEEPINESS AND MENTAL WORKLOAD

There was a significant effect of time measurement on arousal ($F(2, 62) = 13.74$, $p < .001$, $\eta^2 = .07$). Participants reported a higher arousal after both rural ($M = 2.96$, $SE = 0.14$; $t(62) = -4.00$, $p < .001$) and urban ($M = 3.05$, $SE = 0.14$; $t(62) = -4.94$, $p < .001$) environment than before the experiment ($M = 2.56$, $SE = 0.13$). Otherwise, no significant effect of ***Supervision*** or ***Intervention*** were found on arousal, valence and sleepiness ($p > .05$). The effect of the Driving environment was not significant on valence and sleepiness ($p > .05$).

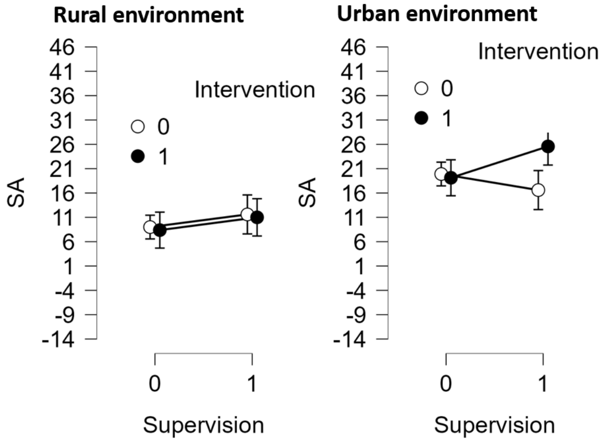


FIGURE 11. Influence of the environment and the AdVitam system (i.e., *Supervision* and *Intervention* modules) on SART results.

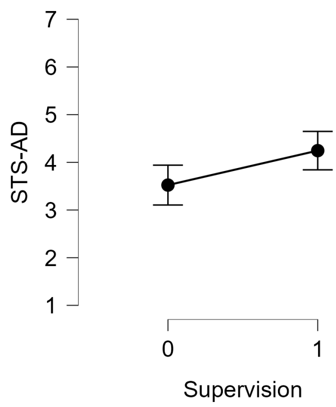


FIGURE 12. Effect of the *Supervision* module on trust.

FIGURE 13. Effect of the *Supervision* module on objective measures of SA (i.e, identification of limitations).

The statistical analysis also revealed a significant effect of the driving environment on workload ($F(1, 31) = 15.62$, $p < .001$, $\eta^2 = .05$). Drivers had a lower mental workload during the rural environment ($M = 15.24$, $SE = 0.60$) than in the urban one ($M = 16.93$, $SE = 0.60$). We can notice that regardless of their condition, drivers reported having a high workload during the experiment. The effect of ***Supervision*** and ***Intervention***, as well as interaction effects were not significant on workload ($p > .05$).

B. SA, TRUST AND USER EXPERIENCE

The statistical analysis revealed a significant effect of the driving environment on SA ($F(1, 31) = 88.92$, $p < .001$, $\eta^2 = .43$). Participants reported having a higher SA in the rural environment ($M = 20.11$, $SE = 0.96$) than in the urban one ($M = 9.80$ $SE = 0.96$) environment. Interestingly, a significant interaction effect of ***Intervention*** and scenario was found on SA ($F(1, 31) = 4.70$, $p < .05$, $\eta^2 = .02$), but post-hoc tests do not help to interpret this result. In the same way, we can observe a significant interaction effect of the ***Supervision***, ***Intervention*** and scenario on SA ($F(1, 31) = 5.00$, $p < .05$, $\eta^2 = .02$), see Figure 11. Besides, the effect of ***Supervision*** and ***Intervention*** modules alone were not significant on SA ($p > .05$).

The statistical analysis also revealed a significant effect of the driving environment on trust ($F(1, 31) = 24.37$, $p < .001$, $\eta^2 = .15$). Participants trusted the car more in the rural environment ($M = 4.36$, $SE = 0.19$) than in the urban one ($M = 3.02$, $SE = 0.19$) environment. Besides, a significant effect of ***Supervision*** was found on trust ($F(1, 31) = 5.32$, $p < .05$, $\eta^2 = .09$). Participants trusted more the car with information provided by ***Supervision*** ($M = 4.24$, $SE = 0.22$) than without ($M = 3.51$, $SE = 0.22$), see Figure 12. Otherwise, the effect of ***Intervention*** was almost significant ($p = .054$), participants tend to trust more the vehicle with ***intervention*** ($M = 4.44$, $SE = 0.24$) than without ($M = 3.61$, $SE = 0.21$). Other all interaction effects were not significant on trust ($p > .05$).

On a scale from -3 to 3, users rated their experience in the simulator at 1.10 ($SD = 0.75$). Pragmatic and hedonic components were respectively rated at 0.66 ($SD = 1.12$) and 1.54 ($SD = 0.77$). Both ***Supervision*** and ***Intervention*** did not statistically affect participants' user experience regarding the global score ($p > .05$). However, we observed a significant effect of ***Supervision*** on the hedonic component of UEQ-S ($F(1, 31) = 8.92$, $p < .01$, $\eta^2 = .21$). Drivers reported to have a higher user experience (in the hedonic component) with ***Supervision*** ($M = 1.89$, $SE = 0.17$) than without ($M = 1.18$, $SE = 0.16$).

C. OBJECTIVES MEASURES OF SA: TASK PERFORMANCE AND IDENTIFICATION OF LIMITATIONS

No effect of driving environment, ***Supervision***, or ***Intervention*** on task performance ($p > .05$). Interaction effects were not significant either ($p > .05$).

The agreement of experimenters for coding the think-aloud answers regarding the Limitations reached 97.57% (12 cases where coding has diverged, over 495 cases). The statistical

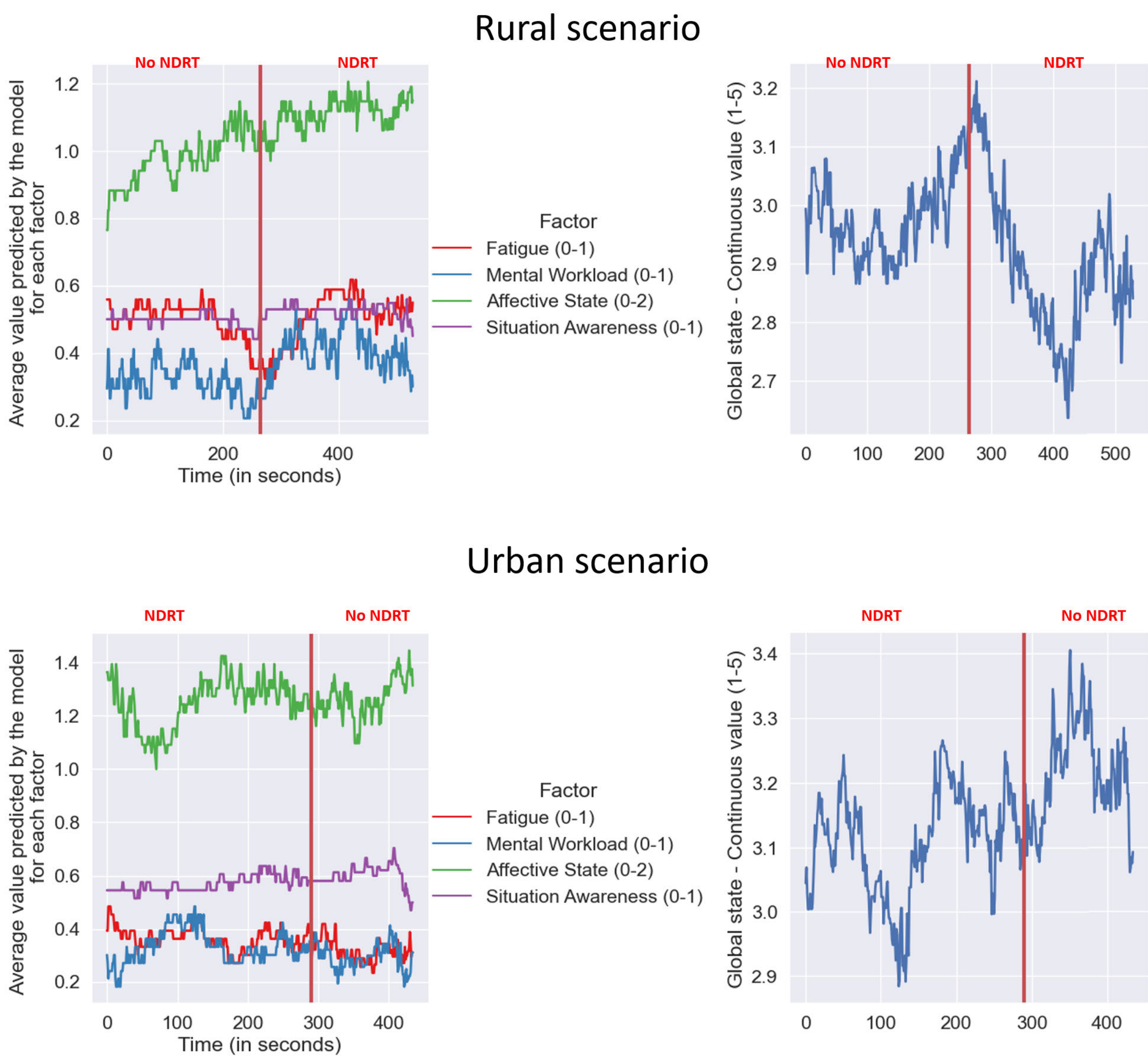


FIGURE 14. **Predictions made continuously by the module 1 (Driver state), for each factor (*Left*) and for the global driver state (*Right*). *Top:* Predictions for the rural scenario. *Bottom:* Predictions for the rural scenario. The horizontal axis is in seconds.**

analysis revealed a significant effect of ***Supervision*** on the identification rate of type ($F(1, 31) = 20.21$, $p < .001$, $\eta^2 = .30$), Severity ($F(1, 31) = 46.69$, $p < .001$, $\eta^2 = .44$) and location ($F(1, 31) = 8.76$, $p < .05$, $\eta^2 = .15$) of limitations occurring during the experiment, see Figure 14. The driving environment, ***Intervention*** and all factor interactions did not have any significant effects on the identification of limitations.

D. TAKEOVER QUALITY

The statistical analysis revealed a significant effect of the driving environment on both reaction time ($F(1, 18) = 28.80$, $p < .001$, $\eta^2 = .37$) and MaxSWA ($F(1, 18) = 5.29$, $p < .05$, $\eta^2 = .12$) after a Take-Over Request. Participants took less time to takeover ($M = 2.38$, $SE = 0.31$) and steered less the wheel ($M = 9.55$, $SE = 9.96$) in the rural environment than in the urban environment ($M = 4.74$, $SE = 0.11$ and $M = 42.66$, $SE = 9.96$). However, no significant effects of ***Supervision*** and ***Intervention*** were observed for either parameter, as well as for all interaction effects between the different factors ($p > .05$).

Fisher's exact tests were used to determine if there was a significant association between the number of crashes and the driving environment, ***Supervision*** and ***Intervention***. There was a statistically significant association between the number of crashes and the driving environment (two-tailed p =.02). There was a significantly higher percentage of crashes in the urban scenario (39.39%) than in the rural one (12.12%). However, there was no association between the number of crashes and ***Supervision*** or ***Intervention*** ($p > .05$).

E. PREDICTIONS MADE BY MODULE 1: DRIVER STATE

Figure 14 shows the prediction of the module for each hazardous factor and for the global state of drivers in both environments.

In the rural environment, participants initially did not engage in the non-driving task on the tablet (first part of the scenario). During this time, Module 1 predicted that their mental workload remained constant (blue line), their fatigue remained constant and then decreased (red line), and their affective state approached a neutral state (green line). Once they began to engage in the cognitive task, the module predicted an increase in their mental workload and fatigue. The predictions made by these blocks were consistent with the prediction of the drivers' overall state. It improved before performing the NDRT but worsened significantly when participants engaged in the cognitive task. Throughout the scenario, the module's prediction of drivers' Situation Awareness remained constant (purple line). For drivers who had the ***Supervision*** module available, predictions of situational awareness ranged from 0.5 to 0.6, whereas they ranged from 0.4 to 0.55 for drivers who did not have this module available (not visible in Figure 14). Overall, we can conclude that the module predictions are consistent with respect to the rural scenario.

In the urban scenario, participants had to engage in NDRT from the beginning. We can see that the module correctly predicted an increase in mental load, then a decrease before remaining constant (blue line). It is difficult to tell if the module made incorrect predictions or if drivers disengaged from the task because it was too difficult to follow in this scenario. The module's predictions for situational awareness (purple line) and fatigue (red line) remained consistent, which makes sense for such a short driving session (about 8 minutes). For the drivers who had the ***Supervision*** module, the predictions of situational awareness increased from 0.6 to 0.75 throughout the scenario, while they remained constant between 0.45 and 0.65 for the drivers who did not have this module (not visible on the figure). For the driver affective state, the predicted value was higher in the urban scenario (about 1.25) than in the rural scenario (about 1). In other words, the module predicts that participants experience more stress in the urban environment, which is consistent with the literature [38].

For the global state predictions in the urban scenario, the drivers' state got worse at the beginning but then improved in the second part of the NDRT. Once they stopped the task, the mental workload slightly decreased and thus the global state improved. However, it is difficult to explain the drastic decrease in drivers' global state at the end of the scenario.

F. FEEDBACKS

Overall, the majority of participants enjoyed the experience (''interesting and immersive experience'', ''a very original experience''), although some had reservations about the operation of the autopilot, particularly in the urban environment (''difficult to know where the vehicle goes in the city''), and sometimes lacked confidence in the system (participants without a module) (''lack of confidence in the system''). The participants confirmed the difficulty of the non-driving related task and the fact that it helped to get them out of the control loop (''difficult to concentrate on several things at the same time''). The majority of participants who had experienced one (or both) of the two modules appreciated receiving this type of information and found it useful (''the device is rather complete and greatly helps for autonomous driving''). Some mentioned that they felt that a little more training would be needed to get the most out of the proposed interactions.

Regarding the takeover request, only one participant preferred having a static TOR (constant modality, without adaptation to the driver state), which indicates a global acceptance of a dynamic TOR. Other feedbacks were mostly about the personal preferences of modalities for the takeover situation, with some participants preferring a voice TOR (''A voice like Siri asking to take over'', ''a message indicating the nature of the problem and its localisation'') while others preferred the auditory chime (''The beep was perfect''). A user suggested always using the auditory modality and adapting only the haptic modality when needed, while another said the auditory modality was not enough for a takeover. All those feedbacks strongly suggest that TOR should be adapted to the user and the situation since no modalities emerged as a global preference.

Regarding the use of the supervision module, participants appreciated the pop-up notifications (''The little notifications on the tablet I was using were great''). Several participants suggested the additional use of chime or speech (''more sound because I am more receptive to it'', ''Sounds in addition to vibrations (e.g. voice saying danger on the right). This would allow looking directly at the right place'').

VIII. DISCUSSION

The results reported above are interpreted in this section. The goal is to validate the effectiveness of the AdVitam system system and the ***Supervision*** and ***Intervention*** Modules on the drivers' state, takeover quality, trust, and user experience.

A. AFFECTIVE STATE, FATIGUE AND MENTAL WORKLOAD

Regarding subjective measures of the driver's state, the data analysis revealed that participants had higher arousal during the driving session than before driving. Conditionally automated driving can hence be considered as an arousing task, even though drivers are not performing the driving task anymore. Participants also reported having a higher mental workload in the urban environment. This effect may be due to the higher number of changing variables in their surroundings (traffic lights and other road users) compared to the monotonous scenario in a rural environment. Overall, participants reported having a rather high mental workload. This suggests that relieving the driver of the driving task does not necessarily reduce the driver's mental load, as the driver may be engaged in another equally cognitive task (as was the case in this experiment). Regarding the module 1 (***Driver State***) predictions, they were often consistent with the experimental manipulation. The predicted fatigue remained constant as the experiment was short, the predicted workload increased during the NDRT (except for the second half in the urban scenario but they might have disengaged for the task), and the affective state was closer to stress in the urban scenario, and the predicted situation awareness was higher for drivers who had the ***Supervision*** module. The predictions of the global driver state were consistent in the rural scenario (worse when engaged in the NDRT) but less in the urban scenario as the predicted driver's state got better after the Take-Over Request. Overall, we can argue that ***(H1)*** is validated. The results obtained are encouraging for further research to continuously assess the driver's state using physiological signals and machine learning.

The use of the ***Supervision*** and ***Intervention*** modules (interaction effect of both) do not have a significant effect on participants' mental workload. This may suggest that receiving information via different interaction modalities combined or adapting dynamically the TOR modality does not add mental load to the driver which is a very important criterion to take into account when creating such interactions. ***(H4)*** is thus verified.

Besides, the nil effects found on sleepiness measures can be explained by the short duration of the experiment. Also, participants were engaged in a non-driving-related task so they did not have time to get drowsy.

B. SITUATION AWARENESS

Based on the analysis of results regarding the identification of Limitations, the ***Supervision*** module has a significant impact on the driver's situation awareness. However, these results are not confirmed by subjective results obtained from the SART questionnaire.

However, the hypothesis that task performance would be higher with the ***Supervision*** module because they would be more engaged in the NDRT while receiving information is not validated. This may be due to the fact that the majority of participants were using this simulator for the first time and were unsure of how the autonomous pilot might behave and therefore preferred to watch the environment more regularly. In addition, some people also mentioned that they found the task very hard and encountered difficulties in performing it. ***(H2)*** is thus partially verified regarding Situation Awareness: the results confirm a positive impact of the module on SA but not on NDRT engagement and performance.

The statistical analysis also shows a significant impact of the type of environment on the driver's situation awareness. Indeed, the results reflect greater SA in rural areas. This can be explained by the fact that the rural scenario proposes few road changes and has no traffic. The situation is then quicker to interpret. The urban environment, on the other hand, offers many more disruptive elements such as traffic and many road changes at intersections.

The results obtained are encouraging for further research to convey adaptive Context-related information according to the situation and driver's state. The use of an interface combinining to propose multisensorial and full-body experience seems beneficial to maintain drivers' SA. The AdVitam system joins the YUI [88] concept which could present similar results.

C. TAKEOVER QUALITY

No beneficial effect of both ***Supervision*** and ***Intervention*** modules were found on takeover quality of participants. ***(H2)*** is thus partially refuted regarding takeover quality and ***(H3)*** is refuted. Getting context-related information did not help drivers to better take over the car. Drivers with the ***Supervision*** module were better at identifying limitations (better SA) but did not show different behaviour in taking over control of the vehicle (no significant difference for reaction time and steering wheel angle). This result is consistent with Endsley's suggestion that having good SA does not necessarily lead to good performance [43]. Given the takeover situations implemented in this study, drivers could have had several possible behaviours to take over control. Significant results could potentially have been observed in a more restricted takeover situation (one choice only, e.g., braking).

Besides, several factors might explain the nil effect of the ***Intervention*** module: the drivers' state was not critical enough in the experiment for the module to have a real impact, the module was trained on too little data, or the module does not take enough Driving environment variables into account. These areas of improvement could increase the impact of the module.

Interestingly, a significant difference in takeover behaviour was found between rural and urban scenarios. The takeover situations were different in both scenarios. In the rural environment, the car requested a takeover because of rocks on the road, while it requested to take over because a dog crossed the road at the last moment in the urban environment. The takeover situation in the rural area yielded lower RT and MaxSWA values. Since both takeover situations were different, results confirm that RT and MaxSWA are relevant metrics to distinguish drivers' takeover behaviour in two different situations. Besides, a higher number of crashes occurred in the urban environment. It seems that the takeover situation was hard to handle for some drivers. This also suggests that 7 seconds before a collision might not be enough to warn the driver for taking over control in an urban environment with traffic while performing a cognitive task.

D. TRUST AND USER EXPERIENCE

The ***Supervision*** module also increased the participants' trust in the automated vehicle. This may fall within the domain of explainable AI. Indeed, participants mentioned appreciating receiving information from the vehicle about what it perceived in the environment. This appreciation is thus well confirmed by the significant results obtained on the STS-AD ratings.

The statistical analysis also shows a significant impact of the type of environment on the driver's trust in the system. Indeed, the results reflect greater trust in rural areas. The urban environment, on the other hand, offers many more lane changing, changing routes at intersections, stopping at red lights, etc... In addition, the participants received no information about the intention of the vehicle which sometimes seemed a bit rough.

Concerning the user experience, the hedonic component is higher with the ***Supervision*** module. This may mean that the experience in the car is more exciting and interesting while receiving notifications. Moreover, the vehicle might also be perceived as more innovative and leading-edge with new interaction modalities that do not exist in conventional cars. ***(H2)*** is thus partially verified for drivers' trust. The ***Supervision*** module shows a positive trend for system acceptance.

A merely significant positive effect of the ***Intervention*** module was found on drivers' trust. Adapting the TOR modality shows a tendency to increase trust in the automated system. This would need to be confirmed in further research as the effect was almost significant here.

No effect was found for the ***Intervention*** module on the user experience in the car. The TOR modality was only seen twice and did not change between the 2 scenarios for some participants. Thus, contrary to the ***Supervision*** module for which the transmitted information changed more often, the participants rather evaluated the theoretical aspect of the ***Intervention*** module than its practical aspect on the user experience. Thus, ***(H3)*** is partially refuted regarding drivers' trust and user experience.

E. LIMITATIONS AND FURTHER WORK

Limitations and prospects for further work are mentioned first for each module, then for the AdVitam system, and finally for the empirical study.

For the ***Driver State*** module, the choice of the four risk factors to be predicted was justified. However, it is questionable and the prediction of other risk factors could be included, such as alcohol or drug use. Furthermore, other approaches for fusion at the level of prediction blocks or reconstruction of the global driver state could be tested. Finally, the continuous predictions of the driver state module were only evaluated qualitatively, but they should also be evaluated quantitatively in future studies. For the ***Supervision*** module, a training phase would have been useful for drivers to better understand the information transmitted by the car. A longitudinal study testing the use of this module at regular intervals could further demonstrate the larger benefits of this user-centred system on drivers' SA and trust. Also, the choice of rules defining the rule-based model was made based on previous experiences evaluating each interface independently [47], [63]. However, the definition of the rules could be modified, after analysing the results of the empirical experiment presented here or other future experiments. For the ***Intervention*** module, it could predict more than 3 combinations of TOR modalities. This would imply collecting more data to train the module.

For the AdVitam system, more variables could be used from the ***Driver State*** module in the predictions made by the ***Supervision*** and ***Intervention*** ones. Also, the robustness has not been measured quantitatively, and this should be done in a further study. Concerning the empirical study, one main limitation is that the AdVitam system has only been tested with young drivers in a static driving simulator. Thus, the study can only be considered a preliminary study for the evaluation of the system. The latter should be tested in real driving conditions and different scenarios of varying duration, involving other takeover situations and other Limitations than erased lane markings or obstacles. It should also be tested with drivers of different age and gender groups, as both factors influence the physiological state and takeover time [89], [90], [91]. Besides, the order of driving scenarios was not randomised, mostly because there were already three experimental factors. This may have induced a learning effect and slightly skew the results. This should be tackled in further experiments.

IX. CONCLUSION

This paper describes the design and implementation of an adaptive system named AdVitam system. The role of this system is to assist the driver in conditionally automated driving and to optimize the shared control between the driver and the vehicle according to the driver's state and the situation. This system is composed of three Modules. The ***Driver State*** module predicts four risk factors (fatigue, mental workload, affective state, situation awareness) as well as the driver's global state. The ***Supervision*** module transmits information related to the context according to a combination of modalities depending on the situation and the state of the driver. Finally, the ***Intervention*** module adapts the modality of control resumption according to the situation and the physiological data of the driver. The design and implementation of the three modules are also described in this paper.

In order to evaluate the AdVitam system, a study with 35 drivers was conducted in a fixed-based driving simulator. Participants experienced two conditionally automated driving scenarios in a rural and urban environment. They encountered different limitations and were tasked with taking over control of the vehicle on demand. A higher mental workload was induced when drivers were engaged in an NDRT during some driving periods. Stress was also manipulated by the type of environment. Measures of driver's state (workload, sleepiness, ...), situation awareness, takeover quality, trust and user experience were assessed.

Firstly, module 1 (***Driver State***) returns a prediction consistent with the driver's experimental manipulation. We observe that the predicted mental workload increases during the NDRT phases (in particular during the rural scenario) and that the predicted affective state is closer to stress during the urban scenario. Moreover, we do not see a major change in fatigue due to the short duration of the experiment. Secondly, receiving contextual information about the car's environment through the multimodal interfaces increases (or at least maintain) Situation Awareness but do not improve takeover quality. We also observe an increase in driver trust and a good acceptance of the ***Supervision*** module. Thirdly, the intelligent adaptation of the Take-Over Request modality (***Intervention*** module) does not improve the takeover quality but tends to increase the driver's trust as it increases the transparency and personalisation of the vehicle. Finally, driving with the full AdVitam system (***Supervision*** $\times$ ***Intervention***) do not increase participants' mental workload and fatigue.

To conclude, this adaptive system increases the driver's SA in automated driving while robustly and continuously assessing the driver's state. Implementing this kind of system could lead to better acceptance of the release of L3-SAE vehicles by increasing driver confidence in autonomous systems.

SUPPLEMENTARY MATERIAL

All the data collected to build the AdVitam system as well as the data from the experiment presented in this article are available on the Zenodo repository: https://doi.org/10.5281/zenodo.7319612.

The attached video illustrates the concept proposed in this article and the experiment performed.

ACKNOWLEDGMENT

(Marine Capallera, Quentin Meteier, and Emmanuel de Salis contributed equally to this work.) The authors would like to thank all the persons who contributed to this paper, especially Renée Favre, Mickael Carvalho, and Arnaud Van Wynsberghe, who participated in implementing the modules.

REFERENCES

[1] *The Top 10 Causes of Death*. World Health Organization, Geneva, Switzerland, 2020.

[2] A. F. Williams and V. I. Shabanova, "Responsibility of drivers, by age and gender, for motor-vehicle crash deaths," *J. Saf. Res.*, vol. 34, no. 5, pp. 527–531, Jan. 2003.

[3] *Testing of Autonomous Vehicles With a Driver*, Dept. Motor Vehicles, CA, USA, 2018.

[4] K. A. Hoff and M. Bashir, "Trust in automation," *Hum. Factors, J. Hum. Factors Ergonom. Soc.*, vol. 57, no. 3, pp. 407–434, Sep. 2015.

[5] O. Jarosch, H. Bellem, and K. Bengler, "Effects of task-induced fatigue in prolonged conditional automated driving," *Hum. Factors, J. Hum. Factors Ergonom. Soc.*, vol. 61, no. 7, pp. 1186–1199, Nov. 2019.

[6] A. W. T. Cai, J. E. Manousakis, B. Singh, J. Kuo, K. J. Jeppe, E. Francis-Pester, B. Shiferaw, C. J. Beatty, S. M. W. Rajaratnam, M. G. Lenné, M. E. Howard, and C. Anderson, "On-road driving impairment following sleep deprivation differs according to age," *Sci. Rep.*, vol. 11, no. 1, p. 21561, Nov. 2021.

[7] B. Mehler, B. Reimer, J. F. Coughlin, and J. A. Dusek, "Impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers," *Transp. Res. Record, J. Transp. Res. Board*, vol. 2138, no. 1, pp. 6–12, Jan. 2009.

[8] G. Matthews, L. Dorn, T. W. Hoyes, D. R. Davies, A. I. Glendon, and R. G. Taylor, "Driver stress and performance on a driving simulator," *Hum. Factors*, vol. 40, no. 1, pp. 136–149, Mar. 1998.

[9] P. Thiffault and J. Bergeron, "Fatigue and individual differences in monotonous simulated driving," *Personality Individual Differences*, vol. 34, no. 1, pp. 159–176, Jan. 2003.

[10] M. Capallera, Q. Meteier, E. de Salis, L. Angelini, S. Carrino, O. A. Khaled, and E. Mugellini, "Owner manuals review and taxonomy of ADAS limitations in partially automated vehicles," in *Proc. 11th Int. Conf. Automot. User Interfaces Interact. Veh. Appl.* New York, NY, USA, 2019. Association for Computing Machinery, Sep. 2019, pp. 156–164.

[11] E. J. D. Ogden and H. Moskowitz, "Effects of alcohol and other drugs on driver performance," *Traffic Injury Prevention*, vol. 5, no. 3, pp. 185–198, Sep. 2004.

[12] S. E. Burian, A. Liguori, and J. H. Robinson, "Effects of alcohol on risk-taking during simulated driving," *Human Psychopharmacology: Clin. Experim.*, vol. 17, no. 3, pp. 141–150, 2002.

[13] R. West, M. John Wilding, D. French, R. Kemp, and A. Irving, "Effect of low and moderate doses of alcohol on driving hazard perception latency and driving speed," *Addiction*, vol. 88, no. 4, pp. 527–532, 1993.

[14] A. Ronen, H. Chassidim, P. Gershon, Y. Parmet, A. Rabinovich, R. Bar-Hamburger, Y. Cassuto, and D. Shinar, "The effect of alcohol, THC and their combination on perceived effects, willingness to drive and performance of driving and non-driving tasks," *Accident Anal. Prevention*, vol. 42, no. 6, pp. 65–1855, 2010.

[15] A. R. Barkley and J. D. Cox, "A review of driving risks and impairments associated with attention-deficit/hyperactivity disorder and the effects of stimulant medication on driving performance," *J. Saf. Res.*, vol. 38, no. 1, pp. 28–113, 2007.

[16] A. Karel Brookhuis, "Driver impairment monitoring by physiological measures," in *Fatigue and Driving*. London, U.K.: Taylor & Francis, 2019.

[17] S. Grégoire Larue, A. Rakotonirainy, and N. A. Pettitt, "Driving performance impairments due to hypovigilance on monotonous roads," *Accident Anal. Prevention*, vol. 43, no. 6, pp. 46–2037, 2011.

[18] G. Maycock, "Driver sleepiness as a factor in car and HGV accidents," Transposition Res. Lab., TRL Rep. 169, 1995.

[19] T. L. Bunn, S. Slavova, and P. J. Rock, "Association between commercial vehicle driver at-fault crashes involving sleepiness/fatigue and proximity to rest areas and truck stops," *Accident Anal. Prevention*, vol. 126, pp. 3–9, May 2019.

[20] T. Akerstedt, A. Anund, J. Axelsson, and G. Kecklund, "Subjective sleepiness is a sensitive indicator of insufficient sleep and impaired waking function," *J. Sleep Res.*, vol. 23, no. 3, pp. 242–254, 2014. [Online]. Available: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jsr.12158

[21] M. Ingre, T. Akerstedt, B. Peters, A. Anund, and G. Kecklund, "Subjective sleepiness, simulated driving performance and blink duration: Examining individual differences," *J. Sleep Res.*, vol. 15, no. 1, pp. 47–53, 2006. [Online]. Available: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2869.2006.00504.x

[22] J. D. McKeever, M. T. Schultheis, V. Padmanaban, and A. Blasco, "Driver performance while texting: Even a little is too much," *Traffic Injury Prevention*, vol. 14, no. 2, pp. 132–137, Jan. 2013.

[23] M. S. Young, K. A. Brookhuis, C. D. Wickens, and P. A. Hancock, "State of science: Mental workload in ergonomics," *Ergonomics*, vol. 58, no. 1, pp. 1–17, Jan. 2015.

[24] P. Thiffault and J. Bergeron, "Monotony of road environment and driver fatigue: A simulator study," *Accident Anal. Prevention*, vol. 35, no. 3, pp. 91–381, 2003.

[25] N. Lyu, L. Xie, C. Wu, Q. Fu, and C. Deng, "Driver's cognitive workload and driving performance under traffic sign information exposure in complex environments: A case study of the highways in China," *Int. J. Environ. Res. Public Health*, vol. 14, no. 2, p. 203, Feb. 2017.

[26] T.-Y. Hu, X. Xie, and J. Li, "Negative or positive? The effect of emotion and mood on risky driving," *Transp. Res. F, Traffic Psychol. Behaviour*, vol. 16, pp. 29–40, Jan. 2013.

[27] L. B. Nabatilan, F. Aghazadeh, A. D. Nimbarte, C. C. Harvey, and S. K. Chowdhury, "Effect of driving experience on visual behavior and driving performance under different driving conditions," *Cognition, Technol. Work*, vol. 14, no. 4, pp. 355–363, Nov. 2012.

[28] A. Guettas, S. Ayad, and O. Kazar, "Driver state monitoring system: A review," in *Proc. 4th Int. Conf. Big Data Internet Things*, 2019, pp. 1–7.

[29] C.-T. Lin, R.-C. Wu, S.-F. Liang, W.-H. Chao, Y.-J. Chen, and T.-P. Jung, "EEG-based drowsiness estimation for safety driving using independent component analysis," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 52, no. 12, pp. 2726–2738, Dec. 2005.

[30] T. Kundinger and A. Riener, "The potential of wrist-worn wearables for driver drowsiness detection: A feasibility analysis," in *Proc. 28th ACM Conf. User Model., Adaptation Personalization*, Jul. 2020, pp. 117–125.

[31] G. Li and W.-Y. Chung, "Detection of driver drowsiness using wavelet analysis of heart rate variability and a support vector machine classifier," *Sensors*, vol. 13, no. 12, pp. 16494–16511, Dec. 2013.

[32] M. A. Hogervorst, A.-M. Brouwer, and J. B. F. van Erp, "Combining and comparing EEG, peripheral physiology and eye-related measures for the assessment of mental workload," *Frontiers Neurosci.*, vol. 8, p. 322, Oct. 2014.

[33] E. T. Solovey, M. Zec, E. A. Garcia Perez, B. Reimer, and B. Mehler, "Classifying driver workload using physiological and driving performance data: Two field studies," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, Apr. 2014, pp. 4057–4066.

[34] Q. Meteier, M. Capallera, S. Ruffieux, L. Angelini, O. A. Khaled, E. Mugellini, M. Widmer, and A. Sonderegger, "Classification of drivers' workload using physiological signals in conditional automation," *Frontiers Psychol.*, vol. 12, Feb. 2021, Art. no. 596038.

[35] Q. Meteier, E. De Salis, M. Capallera, M. Widmer, L. Angelini, O. A. Khaled, A. Sonderegger, and E. Mugellini, "Relevant physiological indicators for assessing workload in conditionally automated driving, through three-class classification and regression," *Frontiers Comput. Sci.*, vol. 3, pp. 1–23, Jan. 2022.

[36] R. W. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: Analysis of affective physiological state," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 10, pp. 1175–1191, Oct. 2001.

[37] J. Kim and E. André, "Emotion recognition based on physiological changes in music listening," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 12, pp. 2067–2083, Feb. 2008.

[38] J. A. Healey and R. W. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 2, pp. 156–166, Jun. 2005.

[39] J.-S. Wang, C.-W. Lin, and Y.-T.-C. Yang, "A K-nearest-neighbor classifier with heart rate variability feature-based transformation algorithm for driving stress recognition," *Neurocomputing*, vol. 116, pp. 136–143, Sep. 2013.

[40] L.-L. Chen, Y. Zhao, P.-F. Ye, J. Zhang, and J.-Z. Zou, "Detecting driving stress in physiological signals based on multimodal feature analysis and kernel classifiers," *Exp. Syst. Appl.*, vol. 85, pp. 279–291, Nov. 2017.

[41] R. Hoeger, A. J. Amditis, M. Kunert, A. Hoess, F. Flemish, H. Krueger, A. Bartels, A. Beutner, and K. Pagle, "Highly automated vehicles for intelligent transport: Haveit approach," in *Proc. ITS World Congr.*, 2008, pp. 1–10.

[42] N. Rauch, A. Kaußner, H.-P. Krüger, S. Boverie, and F. Flemisch, "The importance of driver state assessment within highly automated vehicles," in *Proc. 16th ITS World Congr. Exhib. Intell. Transp. Syst. Services*, 2009, p. 8.

[43] R. Mica Endsley and J. Daniel Garland, *Situation Awareness Analysis and Measurement*. NJ, USA: Lawrence Erlbaum Associates, 2009.

[44] Q. Meteier, M. Capallera, E. De Salis, M. Widmer, L. Angelini, O. Abou Khaled, E. Mugellini, and A. Sonderegger, "Carrying a passenger and relaxation before driving: Classification of young drivers' physiological activation," *Physiological Rep.*, vol. 10, no. 10, May 2022, Art. no. e15229.

[45] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Jan. 2011.

[46] D. Makowski, T. Pham, Z. J. Lau, J. C. Brammer, F. Lespinasse, H. Pham, C. Schölzel, and S. H. A. Chen, "NeuroKit2: A Python toolbox for neurophysiological signal processing," *Behav. Res. Methods*, vol. 53, no. 4, pp. 1689–1696, Aug. 2021.

[47] Q. Meteier, M. Capallera, E. de Salis, A. Sonderegger, L. Angelini, S. Carrino, O. A. Khaled, and E. Mugellini, "The effect of instructions and context-related information about limitations of conditionally automated vehicles on situation awareness," in *Proc. 12th Int. Conf. Automot. User Interfaces Interact. Veh. Appl.* New York, NY, USA: Association for Computing Machinery, Sep. 2020, pp. 241–251.

[48] P. K. Atrey, M. A. Hossain, A. El Saddik, and M. S. Kankanhalli, "Multimodal fusion for multimedia analysis: A survey," *Multimedia Syst.*, vol. 16, no. 6, pp. 345–379, 2010.

[49] M. R. Endsley, "Toward a theory of situation awareness in dynamic systems," *J. Human Factors Ergonomics Soc.*, vol. 37, no. 1, pp. 32–64, Mar. 1995.

[50] C. Bolstad, H. Cuevas, J. Wang-Costello, M. Endsley, and L. Angell, "Measurement of situation awareness for automobile technologies of the future," *SAE Int.*, vol. 11, pp. 195–213, Jan. 2010.

[51] N. Merat, D. B. Seppelt, T. Louw, J. Engström, D. John Lee, E. Johansson, A. C. Green, S. Katazaki, A. C. Monk, M. Itoh, V. D. McGehee, T. Sunda, K. Unoura, T. Victor, A. Schieben, and A. Keinath, "The 'out-of-the-loop' concept in automated driving: Proposed definition, measures and implications," *Cognition, Technol. Work*, vol. 21, pp. 87–98, Sep. 2018.

[52] T. Vogelpohl, M. Kühn, T. Hummel, and M. Vollrath, "Asleep at the automated wheel—Sleepiness and fatigue during highly automated driving," *Accident Anal. Prevention*, vol. 126, pp. 70–84, May 2019.

[53] S. S. Bakker and K. Niemantsverdriet, "The interaction-attention continuum: Considering various levels of human attention in interaction design," *Int. J. Des.*, vol. 10, no. 2, pp. 1–14, 2016.

[54] A. Telpaz, B. Rhindress, I. Zelman, and O. Tsimhoni, "Haptic seat for automated driving: Preparing the driver to take control effectively," in *Proc. 7th Int. Conf. Automot. User Interfaces Interact. Veh. Appl.* New York, NY, USA: Association for Computing Machinery, Sep. 2015, pp. 23–30.

[55] M. Schwalk, N. Kalogerakis, and T. Maier, "Driver support by a vibrotactile seat matrix—Recognition, adequacy and workload of tactile patterns in take-over scenarios during automated driving," *Proc. Manuf.*, vol. 3, pp. 2466–2473, 2015.

[56] A. Löcken, W. Heuten, and S. Boll, "AutoAmbiCar: Using ambient light to inform drivers about intentions of their automated cars," in *Proc. Adjunct 8th Int. Conf. Automot. User Interfaces. Interact. Veh. Appl.* New York, NY, USA: Association for Computing Machinery, Oct. 2016, pp. 57–62.

[57] Y. Yang, B. Karakaya, G. C. Dominioni, K. Kawabe, and K. Bengler, "An HMI concept to improve Driver's visual behavior and situation awareness in automated vehicle," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 650–655.

[58] S. Stockert, N. T. Richardson, and M. Lienkamp, "Driving in an increasingly automated world—Approaches to improve the driver-automation interaction," *Proc. Manuf.*, vol. 3, pp. 2889–2896, 2015.

[59] P. Lindemann, T.-Y. Lee, and G. Rigoll, "An explanatory windshield display interface with augmented reality elements for urban autonomous driving," in *Proc. IEEE Int. Symp. Mixed Augmented Reality Adjunct (ISMAR-Adjunct)*, Oct. 2018, pp. 36–37.

[60] L. Tijerina, M. Blommer, R. Curry, R. Swaminathan, D. S. Kochhar, and W. Talamonti, "An exploratory study of driver response to reduced system confidence notifications in automated driving," *IEEE Trans. Intell. Vehicles*, vol. 1, no. 4, pp. 325–334, Dec. 2016.

[61] G. Wiegand, C. Mai, Y. Liu, and H. Hußmann, "Early take-over preparation in stereoscopic 3D," in *Proc. Adjunct 10th Int. Conf. Automot. User Interfaces Interact. Veh. Appl.* New York, NY, USA: Association for Computing Machinery, Sep. 2018, pp. 142–146.

[62] A. Kunze, S. J. Summerskill, R. Marshall, and A. J. Filtness, "Conveying uncertainties using peripheral awareness displays in the context of automated driving," in *Proc. 11th Int. Conf. Automot. User Interfaces Interact. Veh. Appl.* New York, NY, USA: Association for Computing Machinery, Sep. 2019, pp. 329–341.

[63] M. Capallera, J.-P. Bresciani, Q. Meteier, L. Angelini, O. A. Khaled, and E. Mugellini, "Comparison of human-vehicle interactions concepts to support the driver in conditionally automated driving," in *Proc. 33rd Conf. Interaction Humain-Machine*, 2022, pp. 1–11.

[64] S. Sadeghian Borojeni, S. C. J. Boll, W. Heuten, H. H. Bülthoff, and L. Chuang, "Feel the movement: Real motion influences responses to take-over requests in highly automated vehicles," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2018, pp. 1–13.

[65] I. Politis, S. Brewster, and F. Pollick, "Using multimodal displays to signify critical handovers of control to distracted autonomous car drivers," *Int. J. Mobile Hum. Comput. Interact.*, vol. 9, no. 3, pp. 1–16, Jul. 2017.

[66] Y. Forster, F. Naujoks, A. Neukum, and L. Huestegge, "Driver compliance to take-over requests with different auditory outputs in conditional automation," *Accident Anal. Prevention*, vol. 109, pp. 18–28, Dec. 2017.

[67] S. M. Petermeijer, S. Cieler, and J. C. F. de Winter, "Comparing spatially static and dynamic vibrotactile take-over requests in the driver seat," *Accident Anal. Prevention*, vol. 99, pp. 218–227, Feb. 2017.

[68] A. Telpaz, B. Rhindress, I. Zelman, and O. Tsimhoni, "Using a vibrotactile seat for facilitating the handover of control during automated driving," *Int. J. Mobile Human Comput. Interact.*, vol. 9, no. 3, pp. 17–33, Jul. 2017.

[69] P. Kerschbaum, L. Lorenz, and K. Bengler, "A transforming steering wheel for highly automated cars," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2015, pp. 1287–1292.

[70] S. S. Borojeni, T. Wallbaum, W. Heuten, and S. Boll, "Comparing shape-changing and vibro-tactile steering wheels for take-over requests in highly automated driving," in *Proc. 9th Int. Conf. Automot. User Interfaces Interact. Veh. Appl.*, Sep. 2017, pp. 221–225.

[71] N. Kim, K. Jeong, M. Yang, Y. Oh, and J. Kim, "Are you ready to take-over?": An exploratory study on visual assistance to enhance driver vigilance," in *Proc. CHI Conf. Extended Abstr. Hum. Factors Comput. Syst.*, 2017, pp. 1771–1778.

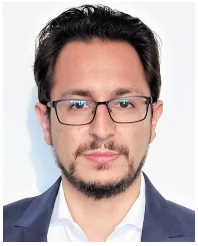
[72] M. Peng, C. Jin, and S. Dihua, "Risk status identification during the takeover of conditionally automated vehicles," in *Proc. Chin. Control Decis. Conf. (CCDC)*, Aug. 2020, pp. 5000–5006.
[73] D. A. Drexler, A. Takacs, T. D. Nagy, and T. Haidegger, "Handover process modeling in autonomous vehicles—A control engineering approach," in *Proc. IEEE 19th Int. Symp. Comput. Intell. Informat. 7th IEEE Int. Conf. Recent Achievements Mechatronics, Autom., Comput. Sci. Robot. (CINTI-MACRo)*, Nov. 2019, pp. 000155–000160.
[74] N. Du, F. Zhou, E. M. Pulver, D. M. Tilbury, L. P. Robert, A. K. Pradhan, and X. J. Yang, "Predicting driver takeover performance in conditionally automated driving," *Accident Anal. Prevention*, vol. 148, Dec. 2020, Art. no. 105748.
[75] C. G. Gold, "Modeling of take-over performance in highly automated vehicle guidance," Ph.D. thesis, Technischen Universität München, Ergonomie, Germany, Munchen, 2016.
[76] H. Kim, W. Kim, J. Kim, S.-J. Lee, and D. Yoon, "A study on the effects of providing situation awareness information for the control authority transition of automated vehicle," in *Proc. Int. Conf. Inf. Commun. Technol. Converg. (ICTC)*, Oct. 2019, pp. 1394–1396.
[77] S. Li, P. Blythe, W. Guo, and A. Namdeo, "Investigation of older driver's takeover performance in highly automated vehicles in adverse weather conditions," *IET Intell. Transp. Syst.*, vol. 12, no. 9, pp. 1157–1165, Nov. 2018.
[78] E. D. Salis, M. Capallera, Q. Meteier, L. Angelini, O. A. Khaled, E. Mugellini, M. Widmer, and S. Carrino, "Designing an AI-companion to support the driver in highly autonomous cars," in *Proc. Int. Conf. Human-Comput. Interact.* Cham, Switzerland: Springer, 2020, pp. 335–349.
[79] E. D. Salis, Q. Meteier, M. Capallera, C. Pelletier, L. Angelini, O. A. Khaled, E. Mugellini, M. Widmer, and S. Carrino, "Predicting takeover quality in conditionally autonomous vehicles based on takeover request modalities, driver physiological state and the environment," in *Proc. Intell. Hum. Syst. Integr. (IHSI) Integrating People Intell. Syst.*, 2022.
[80] E. de Salis, Q. Meteier, M. Capallera, C. Pelletier, L. Angelini, O. Abou Khaled, E. Mugellini, M. Widmer, and S. Carrino, "Machine learning agent to recommend the best modality for takeover during conditionally automated driving," in *Proc. Human Interact. Emerg. Technol. (IHIET-AI), Artif. Intell. Future Appl.*, 2022.
[81] A. Sonderegger, K. Heyden, A. Chavaillaz, and J. Sauer, "AniSAM & AniAvatar: Animated visualizations of affective states," in *Proc. CHI Conf. Hum. Factors Comput. Syst.* New York, NY, USA: Association for Computing Machinery, May 2016, pp. 4828–4837.
[82] T. Akerstedt and M. Gillberg, "Subjective and objective sleepiness in the active individual," *Int. J. Neurosci*, vol. 52, nos. 1–2, pp. 29–37, Jul. 1990.
[83] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (task load index): Results of empirical and theoretical research," *Adv. Psychol.*, vol. 52, pp. 139–183, Apr. 1988.
[84] R. M. Taylor, "Situational awareness rating technique (SART): The development of a tool for aircrew systems design," in *Situational Awareness*. Evanston, IL, USA: Routledge, 2017, pp. 111–128.
[85] B. E. Holthausen, P. Wintersberger, B. N. Walker, and A. Riener, "Situational trust scale for automated driving (STS-AD): Development and initial validation," in *Proc. 12th Int. Conf. Automot. User Interfaces Interact. Veh. Appl.* New York, NY, USA: Association for Computing Machinery, Sep. 2020, pp. 40–47.
[86] Y. Cao, F. Zhou, E. M. Pulver, L. J. Molnar, L. P. Robert, D. M. Tilbury, and X. J. Yang, "Towards standardized metrics for measuring takeover performance in conditionally automated driving: A systematic review," in *Proc. Hum. Factors Ergonom. Soc. Annu. Meeting*, Sep. 2021, vol. 65, no. 1, pp. 1065–1069.
[87] B. Laugwitz, T. Held, and M. Schrepp, "Construction and evaluation of a user experience questionnaire," in *HCI and Usability for Education and Work* (Lecture Notes in Computer Science), A. Holzinger, Ed. Springer, 2008, pp. 63–76.
[88] S. Okamoto and S. Sano, "Anthropomorphic AI agent mediated multimodal interactions in vehicles," in *Proc. 9th Int. Conf. Automot. User Interfaces Interact. Veh. Appl. Adjunct*, New York, NY, USA: Association for Computing Machinery, Sep. 2017, pp. 110–114.
[89] T. B. J. Kuo, T. Lin, C. C. H. Yang, C.-L. Li, C.-F. Chen, and P. Chou, "Effect of aging on gender differences in neural control of heart rate," *Amer. J. Physiol.-Heart Circulatory Physiol.*, vol. 277, no. 6, pp. H2233–H2239, Dec. 1999.
[90] I. A. D. O'Brien, P. O'Hare, and R. J. M. Corrall, "Heart rate variability in healthy subjects: Effect of age and the derivation of normal ranges for tests of autonomic function," *Brit. Heart J.*, vol. 55, p. 348– 354, 1986.
[91] S. Li, P. Blythe, W. Guo, and A. Namdeo, "Investigating the effects of age and disengagement in driving on driver's takeover control performance in highly automated vehicles," *Transp. Planning Technol.*, vol. 42, no. 5, pp. 470–497, Jul. 2019.

MARINE CAPALLERA is currently pursuing the Ph.D. degree with the University of Fribourg and the HES-SO University of Applied Sciences and Arts Western Switzerland. She is also working on conditionally automated driving and focusing more specifically on multimodal human–vehicle interaction model for supervision.

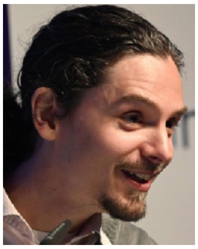
QUENTIN METEIER is currently pursuing the Ph.D. degree with the University of Fribourg and the HES-SO University of Applied Sciences and Arts Western Switzerland. He is also working on conditionally automated driving. His research interest includes evaluating the physiological state of the driver using machine learning techniques.

EMMANUEL DE SALIS is currently pursuing the Ph.D. degree with the University of Fribourg and the HES-SO University of Applied Sciences and Arts Western Switzerland. He is also working on conditionally automated driving. His research interest includes supporting the driver in takeover situations using machine learning techniques.

MARINO WIDMER is a Professor at the University of Fribourg. He teaches operations management and supply chain management and logistics. His research interest includes the development and the application of operations research models to industrial and service problems.



LEONARDO ANGELINI received the Ph.D. degree in computer science conducted in a cooperation between the University of Applied Sciences of Western Switzerland (Fribourg) and the University of Fribourg. He is currently a Professor at the School of Management, Fribourg. He also works as a Researcher at the HumanTech Institute. His research interests include human–computer interaction, with particular focus in gesture interaction, tangible interaction, multimodal interaction, affective computing, with several year of experience in vehicular applications and participative design with seniors.



STEFANO CARRINO received the B.Sc. and M.Sc. degrees in electronic engineering from the University of Florence, Italy, and the Ph.D. degree in computer science realized in a cooperation between the University of Applied Sciences of Western Switzerland, Fribourg, and the University of Fribourg. He is currently a Professor of computer science at the University of Applied Sciences of Western Switzerland in Neuchâtel, (HE-Arc). He has been a member of the Data Analytics Research Group, since 2016. He is also responsible or involved in multiple national and international projects in which data analysis, machine learning, and AI solutions play a crucial role. His main research interest includes the conception of machine learning techniques for their application in several domains from Industry 4.0 to advanced human–computer interaction.

ANDREAS SONDEREGGER is a Professor at the Bern University of Applied Sciences in Business and a Lecturer at the University of Fribourg, where he leads the Human Factors Laboratory. He researches and teaches in the fields of cognitive ergonomics, human–computer interaction, and work and organizational psychology.

OMAR ABOU KHALED received the B.Sc. degree in computer engineering in 1991, and the master's and Ph.D. degrees in computer science from the HEUDIASYC Laboratory, Perception and Automatic Control Group, Université de Technologie de Compiègne. He is currently a Professor with the University of Applied Sciences of Western Switzerland. Since 1996, he has been working as a Research Assistant with the MEDIA Group, Theoretical Computer Science Laboratory, EPFL, in the field of educational technologies and web-based training research field on MEDIT and CLASSROOM 2000 projects.

ELENA MUGELLINI received the master's degree in telecommunication engineering and the Ph.D. degree in telematics and information society from the University of Florence, in 2002 and 2006, respectively. She is currently a Professor with the Information and Communication Department, University of Applied Sciences of Western Switzerland (Fribourg). She is also a Leader of the HumanTech Institute. She is also a member of the Telematics Technology Laboratory, University of Florence. Her current research interests include ambient intelligent, multimodal interaction, natural interface, intelligent data analysis, and information and knowledge management.