

EMERGING TECHNOLOGIES  
**TECHNICAL REPORT**



# Measuring and Predicting Drivers' Takeover Readiness and Supporting Takeover Transitions in Automated Driving

**SEPT 2023**

607 14th Street, NW, Suite 701  
Washington, DC 20005  
202-638-5944  
AAAFoundation.org

© 2023 AAA Foundation for Traffic Safety

## **Title**

---

Measuring and Predicting Drivers' Takeover Readiness and Supporting Takeover Transitions in Automated Driving

*(September 2023)*

## **Authors**

---

Doowon Han<sup>1</sup>, Jundi Liu<sup>1</sup>, Feng Zhou<sup>2</sup>, Dawn Tilbury<sup>1</sup>, Lionel Robert<sup>1</sup>, Lisa Molnar<sup>3</sup>, and X. Jessie Yang<sup>1</sup>

<sup>1</sup>*University of Michigan*

<sup>2</sup>*University of Michigan-Dearborn*

<sup>3</sup>*University of Michigan Transportation Research Institute*

## Foreword

---

With new advanced vehicle technology, drivers can allow automation to take on parts of the driving task. However, when circumstances dictate, drivers must be able to take back control of the vehicle safely and efficiently. Examining ways of supporting these control transitions is imperative and can help to realize the intended safety benefits of vehicle technology.

This technical report summarizes a series of studies and data exercises aimed at examining driver's readiness at taking over control of the vehicle following periods of automated driving as well as support systems that can help guide driver's attention during the takeover. The report should be of interest to researchers, the automobile industry who are working in the domain of advanced vehicle technology, and safety advocates.

C. Y. David Yang, Ph.D.

President and Executive Director  
AAA Foundation for Traffic Safety

## About the Sponsor

---

AAA Foundation for Traffic Safety  
607 14<sup>th</sup> Street, NW, Suite 701  
Washington, D.C. 20005  
202-638-5944

[www.aaafoundation.org](http://www.aaafoundation.org)

Founded in 1947, the AAA Foundation for Traffic Safety in Washington, D.C., is a nonprofit, publicly supported charitable research and education organization dedicated to saving lives by preventing traffic crashes and reducing injuries when crashes occur. Funding for this report was provided by voluntary contributions from AAA/CAA and their affiliated motor clubs, individual members, AAA-affiliated insurance companies, and other organizations or sources.

This publication is distributed by the AAA Foundation for Traffic Safety at no charge, as a public service. It may not be resold or used for commercial purposes without the explicit permission of the foundation. It may, however, be copied in whole or in part and distributed for free via any medium, provided the Foundation is given appropriate credit as the source of the material. The AAA Foundation for Traffic Safety assumes no liability for the use or misuse of any information, opinions, findings, conclusions, or recommendations contained in this report.

If trade or manufacturer's names are mentioned, it is only because they are considered essential to the object of this report and their mention should not be construed as an endorsement. The AAA Foundation for Traffic Safety does not endorse products or manufacturers.

## Table of Contents

---

List of Acronyms .....	vi
Executive Summary.....	1
Introduction .....	3
Predicting Driver Takeover Readiness and Performance .....	3
Supporting Driver Attention During Takeover.....	3
<b>PART 1: Predicting Drivers' Takeover Readiness and Performance Using Machine Learning.....</b>	<b>5</b>
Method.....	5
Participants .....	5
Apparatus and Stimuli .....	6
Experimental Design .....	7
Procedure .....	7
Features for Model Development.....	8
Pre-Processing of Physiological Features.....	9
Fréchet Distance-Based Ground Truth Labels for Takeover Performance .....	9
Generalized Prediction Model .....	11
Data Pre-processing and Train-Test Split .....	11
Model Training and Variable Selection .....	12
Results.....	12
Model Development without PCA.....	13
Model Development with PCA for Variable Selection.....	16
Personalized Model for New Drivers .....	18
Personalized Prediction Model.....	19
Results.....	20
Discussion .....	21
Prediction Models and Measurement Window Length.....	21
Driver Features and Takeover Readiness .....	22
Personalized Models.....	22
Key Takeaways.....	23

<b>PART 2: Supporting Driver Attention During Takeover.....</b>	<b>24</b>
Identifying Potential Hazards During Takeover Events.....	24
Method.....	25
Results and Discussion .....	29
Design and Evaluation of a Gaze Guidance System .....	29
Method.....	30
Results.....	33
Discussion.....	37
<b>General Discussion .....</b>	<b>38</b>
<b>References .....</b>	<b>39</b>
<b>Appendix A: Coding Rules for Hazard Alert Conditions .....</b>	<b>44</b>

## List of Acronyms

---

ADAS	Advanced Driver Assistance System
AV	Automated Vehicle
BSD	Blind Spot Detection
CSW	Curve Speed Warning
DSM	Driver State Monitoring Systems
DT	Decision Tree
FCW	Forward Collisions Warning
FD	Fréchet Distance
GSR	Galvanic Skin Response
IVBSS	Integrated Vehicle-Based Safety Systems
LCM	Lane Change/Merge Warning
LDW	Lane Drift Warning
LR	Linear Regression
ML	Machine Learning
NDRT	Non-Driving Related Tasks
PCA	Principal Component Analysis
PPG	Photoplethysmogram
RF	Random Forest
SAE	Society for Automotive Engineers
TOR	Takeover Request

## Executive Summary

---

As vehicle automation progresses, the driver's role will transform from an operator to a system supervisor. With higher levels of automation, the automated vehicle (AV) is able to monitor the environment, which allows the driver to engage in non-driving related tasks (NDRTs). However, if the automated vehicle reaches its system limit (e.g., automation failure, adverse weather, lane markings disappear), the driver will be forced to resume control of the vehicle in a limited amount of time. Unfortunately, when drivers are decoupled from the operational level of control, they often have difficulty taking over in any situation, particularly in situations that the automation is not able to handle.

This report, presented into two parts, highlights two studies aimed at facilitating takeover transitions in Level 3 automation. Part 1 examines driver takeover readiness; that is, driver behavior and physiological indices and other factors that are predictive of successful takeover performance. Knowledge of such measures can inform the development and tuning of driver state monitoring (DSM) systems. Part 2 examines a driver support system, a gaze guidance system, that helps orient drivers' attention to areas of potential risk during a control takeover. This study leverages data from an existing naturalistic driving study as well as theoretical models of driver visual attention allocation.

With respect to predicting driver takeover readiness and performance, a human-subject experiment was conducted and a variety of machine learning (ML) models were developed and tested. In a driving simulator study, 32 participants were requested to take over control from automated driving while playing a video game on a tablet. Drivers' physiological data, including heart rate indices, galvanic skin response (GSR) indices, and eye-tracking metrics, were collected and used as inputs to the ML models. Different modeling approaches were explored and tested to offer insights into (a) what modeling approach was best suited for the data; (b) what is the appropriate measurement window (time) for the different indices; (c) what measures or indices were most important in characterizing driver takeover readiness; and (d) the value of using personalized models (versus using general ones). Key highlights from these efforts include the following:

- In examining driver physiological data in real-time, a random forest (RF) machine learning approach led to the best model predictions. Driver state monitoring systems (and their underlying algorithms) should consider such approaches.
- Pre-takeover measurement windows of 9 to 14 seconds showed the highest model performance, with peak performance for the RF model occurring at 11 seconds. DSM systems should strive to incorporate and/or validate their own outcomes using such time frames.

- GSR indices were the most important measures in predicting driver performance, followed by heart rate indices. As indicators of driver takeover readiness, these measures should be considered for inclusion in DSM systems as part of a suite of measures.
- Personalized models have great potential for increasing model accuracy in real-world implementations. Generalized models that learn from new users have the potential to increase their accuracy and real-world utility.
- The novel measure of takeover performance, based on the Fréchet distance, considers both temporal and spatial characteristics and is therefore useful in capturing the timeliness and quality of takeover performance.

For the second part, supporting driver attention during takeover, a study was conducted to explore avenues to support driver attention during takeover events. It comprised two pieces: (a) identification of potential hazards that co-occur during driver takeover transitions and (b) the design and evaluation of a gaze guidance system to support drivers' noticing of potential hazards during takeovers. The first exercise utilized naturalistic driving data from an existing database to identify types of potential hazards identified using a grounded approach in situations where advanced driver assistance systems (ADAS) issued alerts to the driver related to nearby hazards. The second exercise comprised a driving simulator study aimed at designing and evaluating a gaze guidance system, based on a theoretical model of selective visual attention. A driving simulator study was conducted wherein drivers were exposed to various takeover scenarios with the support of one of two types of gaze guidance system (high and low salience). The results showed that drivers using a highly salient attention guidance system were less likely to become involved in a collision with a secondary hazard during takeover transitions. The results suggest that gaze guidance (attentional) support is a viable approach to helping drivers during takeover events and worthy of further research and innovation.

## Introduction

---

As vehicle automation progresses, the driver's role will transform from an operator to a system supervisor. Level 3 automated vehicles (AVs) possess the ability to perceive their surroundings and interpret road conditions while performing driving tasks such as accelerating, braking, steering, and navigating. The advanced capability allows the driver to engage in non-driving related tasks (NDRTs). However, if an AV encounters a system limit, such as vision system failure or path planning issues, the driver must quickly regain control of the vehicle. This transition from automated control to manual control presents a crucial challenge to the human driver, as they become increasingly out of the loop (OOTL) (Zhou et al., 2020; Petersen et al., 2019; Molnar et al., 2017)

To help inform and address these issues, this report is organized into two parts, presenting two studies aimed at facilitating takeover transitions when using Level 3 automation. Part 1 examines driver takeover readiness; that is, driver behavior and physiological indices and other factors that are predictive of successful takeover performance. Knowledge of such measures can inform the development and tuning of driver state monitoring (DSM) systems. Part 2 examines a driver support system, a gaze guidance system, that helps orient driver's attention to areas of potential risk during a control takeover. This study leverages data from an existing naturalistic driving study as well as theoretical models of driver visual attention allocation.

### **Predicting Driver Takeover Readiness and Performance**

Part 1 presents a human-subject experiment and a variety of machine learning (ML) models for predicting drivers' takeover readiness and performance. A human-in-the-loop experiment was conducted with 32 participants, wherein drivers were requested to take over control from automated driving while playing a Tetris game. Drivers' physiological data, including heart rate indices, galvanic skin response (GSR) indices, and eye-tracking metrics, were collected and used as inputs to the ML models. Different modeling approaches were explored and tested to offer insights into (a) what ML approach was best suited for the data; (b) what is the appropriate measurement window (time) for the different indices; (c) what measures or indices were most important in characterizing driver takeover readiness; and (d) the value of using personalized models (versus using general ones).

### **Supporting Driver Attention During Takeover**

Part 2 sought to explore avenues to support driver attention during takeover events. It comprises two pieces: (a) identification of potential hazards that co-occur during driver takeover transitions and (b) the design and evaluation of a gaze guidance system to support drivers' noticing of potential hazards during takeovers. The first exercise utilized naturalistic driving data from an existing database, the Integrated Vehicle-Based

Safety System (IVBSS) program (Sayer et al., 2011). Ten types of potential hazards were identified using a grounded approach in situations where ADAS systems issued alerts to the driver related to nearby hazards. Researchers documented the co-occurrence of other nearby hazards (not ones that triggered the alert) that could conceivably impact driver responses to the alert or situation. The second exercise comprised a driving simulator study aimed at designing and evaluating a gaze guidance system, based on the N-SEEV (Salience, Effort, Expectancy, Value) model of visual attention (Wickens, 2015). A human-in-the-loop experiment was conducted wherein drivers were exposed to various takeover scenarios with the support of one of two types of gaze guidance system (high and low salience).

## **PART 1: Predicting Drivers' Takeover Readiness and Performance Using Machine Learning**

---

Existing studies have examined factors that influence drivers' takeover performance. A wide range of factors have been identified, including drivers' characteristics (Clark and Feng, 2017; Du et al., 2020b; Wan and Wu, 2018; Zeeb et al., 2017), driving and external environments (Gold et al., 2016; Li et al., 2018), and the design of driver-vehicle interface (Eriksson et al., 2018; Helldin et al., 2013). These empirical studies have documented the considerable variability in taking over behavior and control in different driving environments and situations.

These studies shed light on the relationships between certain factors and takeover performance; for instance, high traffic density harmed takeover performance (Gold et al., 2016). However, with few exceptions (e.g., Gold et al., 2018; Braunagel et al., 2017; Du et al., 2020c), little effort has been made to integrate these findings into computational models capable of predicting drivers' takeover performance in real time. Moreover, research to date has largely focused on predicting either takeover timeliness or quality. There is a gap in quantifying both temporal and spatial characteristics of the takeover readiness. That is, examining upstream indices occurring prior to the takeover request in order to identify cues that are predictive of driver readiness. Knowledge of such factors can inform the development and implementation of DSM systems.

Therefore, this project aims to measure and model factors that predict drivers' takeover readiness by analyzing their physiological data. A human-in-the-loop experiment with 32 participants was conducted, wherein drivers were requested to take over control from automated driving while playing a Tetris game. Drivers' physiological data, including heart rate indices, GSR indices, and eye-tracking metrics, were collected and takeover quality was assessed using a novel metric to quantify performance. Different ML modeling approaches were explored and tested to offer insights into the ideal modeling approach and parameters as well as the ideal measurements for characterizing driver takeover readiness.

### **Method**

#### ***Participants***

A total of 32 participants (average age = 25.8 years, SD = 4.4 years, 16 females, 16 males) with normal or corrected-to-normal vision participated in the experiment. Each participant received a payment of 30 dollars for their participation. The study was approved by the University of Michigan Institutional Review Board.

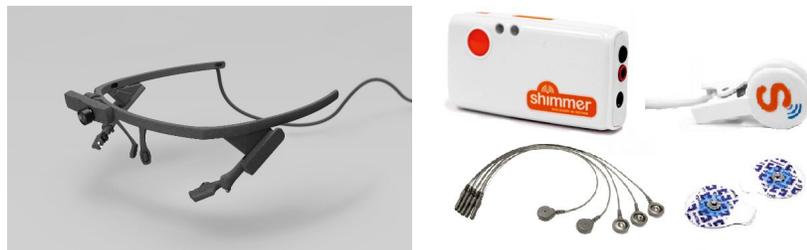
## *Apparatus and Stimuli*

**Driving Simulator and Automated System.** The study was conducted in a fixed-base driving simulator from Realtime Technologies Inc. (RTI, Michigan). The virtual world was displayed on three monitors, located approximately 4 feet in front of the participant (See Figure 1). The simulated vehicle was controlled by a steering wheel and pedal system. The vehicle was programmed to simulate the behavior of an SAE Level 3 automation, which handled the longitudinal and lateral control and navigation, and responded to traffic elements. Participants could press the button on the steering wheel to activate the automated mode, which was indicated by an auditory sound and a blue light on the dashboard. Whenever the AV reached its system limit, a take-over request was issued, consisting of a spoken auditory warning (“Takeover”) and the disappearance of the blue light on the dashboard. At the same time, the automated mode would be automatically deactivated for the driver to take over control of the vehicle.



*Figure 1. RTI driving simulator (left) and non-driving related task (Tetris) (center, right).*

**Physiological Sensors.** During the experiment, participants wore the Pupil Core Glasses eye-tracking system (Pupil Labs, Germany) that provided real-time gaze and pupil data (Figure 2). The sampling rate of the eye-tracking system was 200 Hz. The Shimmer3 GSR+ unit (Shimmer, MA, USA), including GSR electrodes and photoplethysmogram (PPG) probe, was used to collect participants’ GSR and heart rate data with a sampling rate of 128 Hz (Figure 2). The iMotions software (iMotions, MA, USA) was used for physiological data synchronization.



*Figure 2. Physiological sensors: eye tracker (left) and Shimmer GSR and heart rate monitor (right).*

**Non-Driving Related Task.** Participants were asked to play a Tetris game when the vehicle was in automation mode. The NDRT task was selected to promote increased drivers’ eyes-off-the-road and hands-off-the-wheel conditions anticipated in SAE Level 3

automated driving mode. In a Tetris game, puzzle tiles were randomly generated and presented. The task was running on an 11.6-inch touchscreen tablet mounted next to the steering wheel (see Figure 1).

### ***Experimental Design***

The experiment used a 2 × 2 within-subjects design with two independent variables: takeover lead time and tablet location. The two variables were chosen to induce a varying degree of takeover performance. Takeover lead time has been shown to significantly influence takeover performance (Eriksson and Stanton, 2017; Du et al., 2020a), and tablet location also impacts drivers’ car following performance (Lamble et al., 1999). Benchmarking previous studies (Eriksson and Stanton, 2017; Du et al., 2020a), the takeover lead time was set to 4 seconds or 7 seconds. The tablet was placed 25 or 55 centimeters away from the steering wheel for the participant to perform the secondary task. The sequence of the four conditions was balanced using a balanced Latin square design. Each participant experienced four takeover events in the experiment, as described in Table 1.

*Table 1. Description of takeover events.*

<b>Event</b>	<b>Scenario Type</b>	<b>Description</b>
1	Construction Ahead	Urban, two-lane road—construction blocking the lane (traffic cones and equipment) requiring that driver change lanes to the adjacent (vacant) lane.
2	Police Vehicle on Shoulder	Rural two-lane road—police parked on side of road partially obstructing the righthand lane, requiring driver to change lanes.
3	Stationary Bus Ahead	Urban two-lane road—a bus is stopped unexpectedly in the lane ahead, requiring that the driver change lanes.
4	Lead Vehicle Sudden Stop	Rural two-lane road—a lead vehicle brakes abruptly, forcing the driver to brake and/or change lanes.

### ***Procedure***

Upon arrival, participants provided informed consent and filled out a demographic survey. The participants were given an introduction that described the experiment. Next, participants received a training session to familiarize themselves with the driving simulator and the Tetris game. During the training, participants practiced how to drive, change lanes, and engage/disengage the automated driving mode. They were also introduced to the visual display and auditory alerts for takeover requests. Participants were then asked to drive the simulator until they felt comfortable handling the simulator controls. After that, participants played the Tetris game to get familiar with it.

After the training, participants were fitted with the eye tracker. In addition, experimenters attached two GSR electrodes to the participants' left foot and the PPG probe to their left ear lobe. After sensor location adjustment and eye-tracking calibration, participants started the main experiment.

As shown in Figure 3, the main experiment began with the command to activate the automated driving mode. After that, there was an NDRT phase where participants were asked to play the Tetris game. Participants were informed that there was no need to monitor the environment when the AV was in automated driving mode. Once a takeover request (TOR) was issued, participants were required to take over control of the vehicle immediately. Participants were instructed to comply with all the traffic laws when they drove manually. They were informed that the speed limit was 35 mph. Participants could hand back the control to the AV after they negotiated the driving situation. When the AV was re-engaged, participants were asked to complete a short verbal questionnaire.

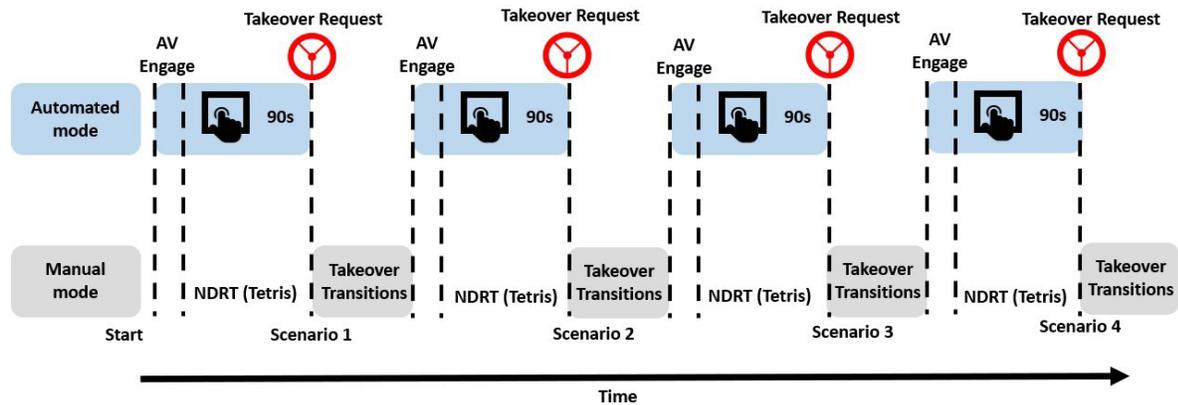


Figure 3. Sequence of takeover events in the experiment.

## Features for Model Development

Drivers' physiological and driving dynamics data were gathered in the study. The physiological data, including heart rate-related features, GSR-related features, and eye-tracking features, were used as inputs for the model development for evaluating takeover readiness. Detailed descriptions of the physiological features are presented in Table 2. The driving dynamics data were processed to calculate the proposed ground truth labels of the takeover performance (i.e., the quality of takeover performance). In total, 32 participants' data were processed and used to train and test the proposed prediction models.

Table 2. Descriptions of physiological and condition-related features.

Feature	Description
HR Indices	Mean, min, max, and standard deviation of heart rate, inter-beat interval
GSR Indices	Mean, max, and standard deviation of GSR in phasic component
GSR Peak	Number of GSR peaks, and peak rise time
Fixation	Fixation number and duration in different areas of interests (i.e., driving scenes and NDRT tablet)
Blink	Number of blinks
Gaze Dispersion	Standard deviation of the values for gaze angle from the right front (radians)
Eyes-on-Road	Proportion of time that participant's gaze is on the road
TOR Lead Time	Short (4s) or long (7s) TOR lead time

Note: HR = heart rate; min = minimum; max = maximum; GSR = galvanic skin responses; TOR = takeover request.

### ***Pre-Processing of Physiological Features***

During the data pre-processing, 36 features, including 28 heart rate-related and GSR-related indices, 7 eye-tracking indices, and 1 environmental feature, takeover lead time, were extracted. Heart rate indices were calculated for mean, minimum, maximum, and standard deviation. For gaze behaviors, drivers' blink number, horizontal gaze dispersion, fixation of road or NDRTs, and fixation durations were calculated. GSR phasic components were extracted from raw GSR signals using the continuous decomposition analysis via Ledalab in Matlab (Benedek & Kaernbach, 2010). Maximum and mean GSR phasic activation were calculated to indicate drivers' arousal and stress in response to TORs (Wintersberger et al., 2018). For eye-tracking features, horizontal gaze dispersion was defined as the standard deviation of gaze heading. Decreases in horizontal gaze dispersion and blink number indicated increases in cognitive load and decreases in attention allocation (Wang et al., 2014; Merat et al., 2012). All features are listed in Table 2.

### ***Fréchet Distance-Based Ground Truth Labels for Takeover Performance***

A novel metric was used to quantify takeover performance based on the Fréchet Distance (FD) (Eiter & Mannila, 1994; Alt & Godau, 1995) between the theoretical optimal takeover trajectory and the actual takeover trajectory. Based on this metric, the more closely the actual takeover trajectory resembles the theoretical optimal, the better the takeover performance.

The theoretical optimal trajectory is the solution to minimize a cost function provided in Abbas et al. (2017). Its goal is to achieve a tradeoff between tracking the

center line of the corridor and regulating the steering rate input while avoiding obstacles. In addition, as Febbo et al. (2017) stated, the way to avoid obstacles in Abbas et al. (2017), namely the soft constraint method, cannot be guaranteed to avoid collision in all cases. Therefore, this work adopted the hard constraint method. Moreover, the trajectory planner changed the lane when the headway to the obstacle is 4 or 7 seconds ahead (Happee et al., 2017). Then CasADi (Andersson et al., 2019) was implemented to transfer the optimal control problem into a nonlinear optimal control problem, which was then solved by the nonlinear program solver IPOPT (i.e., interior point optimizer; Wächter & Biegler, 2006). The open loop solution was then used as the theoretical optimal trajectory.

The trajectories contained temporal and spatial information of the optimal vehicle states at different time points. After obtaining the theoretical optimal trajectory, the FD between the theoretical optimal and the actual trajectories was calculated using the following equation:

$$FD(A, B) = \inf_{\alpha, \beta} \max_t \{d(A(\alpha(t)), B(\beta(t)))\} \quad (1)$$

where  $d$  is the euclidean distance in  $\mathbb{R}^2$ ,  $\alpha$  and  $\beta$  are non-decreasing surjective functions, and  $t$  denotes the time point between TOR and the end of the takeover maneuver.

FD is widely used to measure the similarity between two curves, which considers the location and ordering of the points along the curves. Successful applications included route identification (Lyu et al., 2021) and driver-car matching (Meng et al., 2019).

**Visualization of Optimal and Actual Trajectories.** Figure 4 illustrates the theoretical optimal and the actual trajectory for a sample takeover event. The orange line denotes the theoretical optimal, and the blue line indicates the actual trajectory. The X-axis and Y-axis denote the lateral and longitudinal coordinates of the simulation world, respectively. Compared to the theoretical optimal, the actual driving trajectory in this example revealed a longer response time after receiving the TOR. Moreover, lateral control of the vehicle was initially unstable after the transition of control but eventually stabilized.

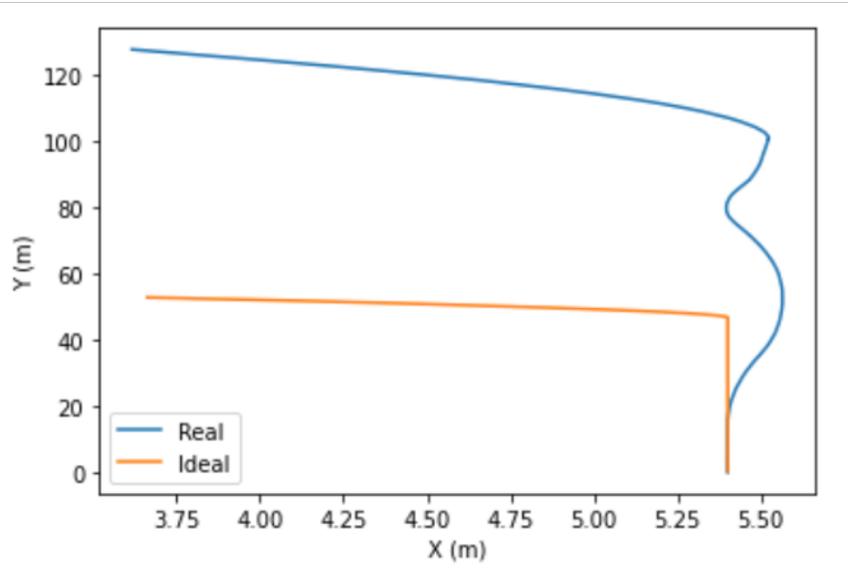


Figure 4. Illustration of the theoretical optimal and the actual driving trajectory for a takeover event in the simulation world. The orange line denotes the theoretical optimal; the blue line denotes the actual trajectory.

## Generalized Prediction Model

### *Data Pre-processing and Train-Test Split*

Each participant experienced four takeover events, so data from 128 takeovers were used for the modeling. For each takeover event, the driving dynamics data were processed to generate the FD-based ground truth labels for the takeover performance. Furthermore, the physiological features for each takeover event were processed using a moving window method to augment the training and testing datasets to a larger sample size. The moving or sliding window method is common for handling time-series data. Since the participants were in a steady mental state and engaged in the secondary task without any interrupting stimuli before the TOR, it was assumed that the physiological features were identically distributed within a short time horizon immediately before the TOR, thus allowing the moving window technique to be leveraged to augment the datasets. The time horizon of interest was defined as the 20 second window prior to the TOR. The size of the moving window was varied from 1 second to 20 seconds and the step length was fixed to 1 second. For example, if the window length is 5 seconds and the step length is 1 second, the first window is 20 seconds to 15 seconds before the TOR, and the second window is 19 seconds to 14 seconds before the TOR, etc. Twenty-eight physiological features and seven eye-tracking features were calculated within each window and attached the same label from the specific takeover event according to the assumptions. All features were normalized to values between 0 and 1 using the MinMax Scaler method. Finally, the data were randomly split 80% into the training set and 20% into the testing set.

## ***Model Training and Variable Selection***

The conventional training and testing method of ML models was applied to compare the performance of five different ML models, including Linear Regression (LR), Ridge Regression (Ridge), Lasso Regression (Lasso), Decision Tree (DT), and Random Forest (RF), for different time window lengths before the TOR. This step aimed to identify the optimal ML model for predicting takeover performance and the optimal window length. All 36 predictors were included for model development: GSR-related, heart rate-related, eye-tracking, and takeover lead time.

Different moving window lengths were compared for sampling the physiological features before the TOR to identify the optimal time window length for predicting takeover performance. The model performance was evaluated using Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Absolute Error (RMSE), and adjusted training and testing  $R^2$  values. The adjusted  $R^2$  values are significant metrics in evaluating the model adequacy in terms of the proportion of the variability explained by the model. The findings have implications on the design of the prediction model in practical applications, shedding insights into the appropriate time window and frequency in collecting and calculating the physiological features as input of the model.

Considering that human subject data are difficult to collect in practical applications, the large number of available features in the current experimental design is an ideal case; however, such a scenario may not be realistic when the resources to collect all 36 features are limited. Therefore, the most significant groups of features in predicting the takeover performance were identified and the lower bounds of model performance using only limited features were explored. Principal Component Analysis (PCA) was applied on the input features to reduce the dimensionality of the dataset to keep the partial variability in the original dataset in order to test the performance under constraints and gain insights into the relative importance of different features (measures). The principal components (PCs) are the linear combinations of the input features, which retain the variables with the most variability. By examining the loadings of each PC, one can identify the most significant variables in the prediction models to provide better explainability of the prediction models.

## **Results**

All model development and testing was carried out on a MacBook Pro with M1 Pro chip and 16 GB RAM running MacOS Ventura 13.1. Python 3.9 and packages including pandas (McKinney, 2010), numpy (Harris et al., 2020), scikit-learn (Pedregosa et al., 2011), and other supporting packages were used for model development and testing. The optimal hyperparameters were trained using the GridSearch method provided by the scikit-learn package.

### Model Development without PCA

Figure 5 and Table 3. Testing  $R^2$  values of the five proposed models of different window lengths without dimension reduction. Table 3 shows the average adjusted testing  $R^2$  values of the five ML models using different moving window lengths without dimension reduction. The RF model outperformed the other four models for different moving window lengths. With the exception of moving window lengths of 19 and 20 seconds, the average adjusted  $R^2$  values of the rest window lengths were over 0.7, meaning the RF model captured approximately 70% of the variability on average for window lengths between 1 to 18 seconds. The DT model was the second-best prediction model. The other three models, LR, Ridge, and Lasso, significantly underperformed compared to the RF and DT models. The results suggest that the input features have a nonlinear relationship to the outcomes since the RF model is an ensemble model, which can capture the underlying nonlinear trend in the dataset. In contrast, the three linear models failed to capture the nonlinear trend resulting in poor prediction performance. Finally, from Table 3, the optimal window length for the RF model in predicting the takeover performance was 11 seconds with an adjusted testing  $R^2$  value of 0.94. There was an overall trend with prediction performance increasing from the 1-second window length to the 11-second window length and decreasing from the 11-second window length to the 20-second window length. In general, however, the model performed highly in the 9-to-15-second range.

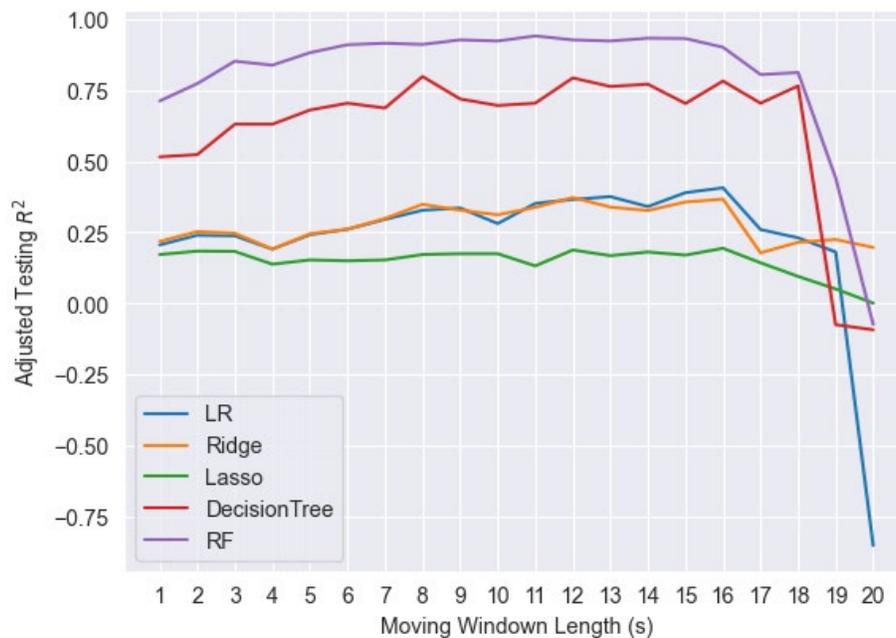


Figure 5. Adjusted testing  $R^2$  values for the five ML models of moving window lengths from 1 second to 20 seconds.

Table 3. Testing  $R^2$  values of the five proposed models of different window lengths without dimension reduction.

Window Length	LR	Ridge	Lasso	DT	RF
1	0.206	0.218	0.172	0.5	0.713
2	0.24	0.253	0.184	0.529	0.774
3	0.238	0.247	0.183	0.627	0.853
4	0.191	0.191	0.138	0.615	0.839
5	0.242	0.245	0.153	0.678	0.883
6	0.261	0.261	0.15	0.693	0.911
7	0.296	0.299	0.153	0.674	0.916
8	0.328	0.349	0.172	0.807	0.912
9	0.336	0.328	0.175	0.705	0.928
10	0.281	0.312	0.175	0.658	0.924
11	0.352	0.338	0.132	0.707	0.942
12	0.366	0.373	0.188	0.789	0.928
13	0.376	0.339	0.168	0.758	0.924
14	0.341	0.327	0.181	0.788	0.934
15	0.39	0.357	0.17	0.76	0.933
16	0.407	0.367	0.194	0.793	0.902
17	0.26	0.178	0.143	0.777	0.806
18	0.231	0.215	0.095	0.763	0.813
19	0.181	0.225	0.051	-0.107	0.442
20	-0.853	0.197	0.001	-0.093	-0.074

The average adjusted training  $R^2$  values were also plotted for the five ML models using different moving window lengths without dimension reduction. In Figure 6, a similar trend is observed in the training values as for the testing  $R^2$  values. The RF model outperformed the other four models for different moving window lengths except for the 16-second and 17-second windows, where the DT model performed better. However, as shown in Table 3, these two window lengths do not achieve better testing performance for the DT model compared to the RF model. Furthermore, the other three linear models still significantly underperform compared to the DR and RF models. Finally, from Table 4, all training  $R^2$  values were above 0.85, which suggests that the proposed models can capture more than 85% of the total variability in the training set. Among all different training  $R^2$  values, the highest values were generally observed for the 9-to-14-second window length.

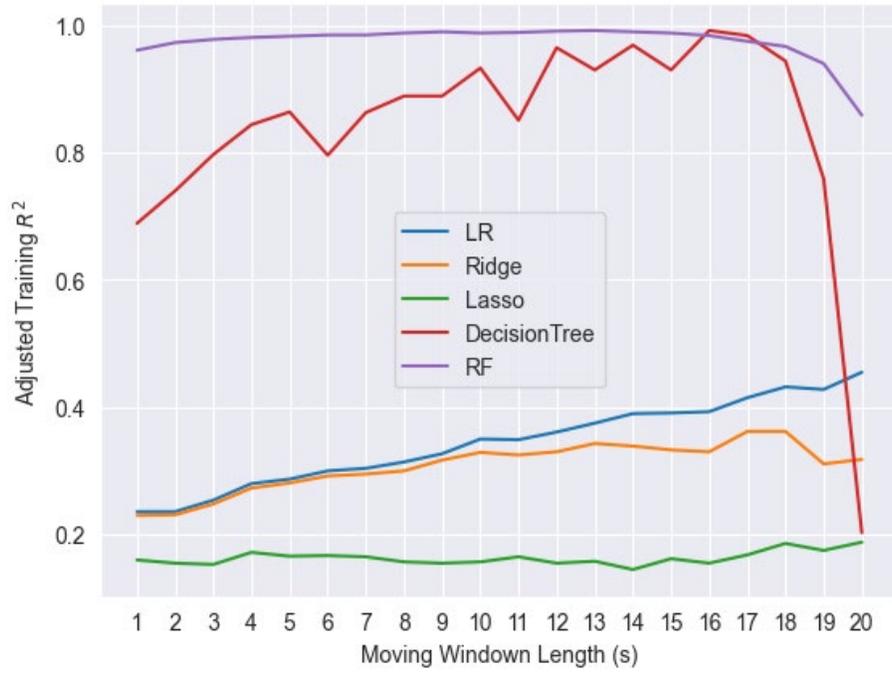


Figure 6. Adjusted training  $R^2$  values for the five ML models of moving window lengths from 1 second to 20 seconds.

Table 4. Training  $R^2$  values of the five proposed models of different window lengths without dimension reduction.

Window Length	LR	Ridge	Lasso	DT	RF
1	0.236	0.23	0.16	0.689	0.961
2	0.236	0.231	0.155	0.739	0.973
3	0.254	0.248	0.153	0.814	0.978
4	0.28	0.273	0.172	0.844	0.981
5	0.287	0.281	0.166	0.864	0.983
6	0.3	0.292	0.167	0.796	0.985
7	0.304	0.295	0.165	0.855	0.985
8	0.314	0.3	0.157	0.889	0.988
9	0.327	0.317	0.155	0.889	0.99
10	0.35	0.329	0.157	0.932	0.988
11	0.349	0.325	0.165	0.853	0.989
12	0.361	0.33	0.155	0.965	0.991
13	0.375	0.343	0.158	0.935	0.992
14	0.39	0.339	0.145	0.969	0.99
15	0.391	0.333	0.162	0.93	0.988
16	0.393	0.33	0.155	0.992	0.984
17	0.415	0.362	0.168	0.984	0.975
18	0.432	0.362	0.186	0.944	0.967
19	0.428	0.311	0.175	0.826	0.94
20	0.455	0.318	0.188	0.203	0.859

However, the relatively smaller testing  $R^2$  values observed for the 9-to-10- and 12-to-14-second time windows suggests overfitting of data in the training set. Therefore, the 11-second window length was selected for subsequent modeling.

### ***Model Development with PCA for Variable Selection***

Human subject data are often difficult to collect for a large sample size and for a large number of measures or features. In practical applications, there might only be access or budget for certain groups of features based on the available sensors. In the previous section, the training and testing  $R^2$  values indicated that the RF model performed very well in many instances, especially for an 11-second time window. To test the model performance under imperfect or incomplete observations and gain insights into the relative significance of the variables, a PCA was applied to reduce the dimensions and maintain partial variability of the input data to select the most significant input features, using the 11-second moving window length.

To select the optimal number of PCs, the scree plot of the first five PCs with the largest proportions of explained variance is plotted in Figure 7. Although the first elbow point appears at the second PC, the second elbow point at the fourth PC was selected in order to maintain over 70% of the total variability in the training set.

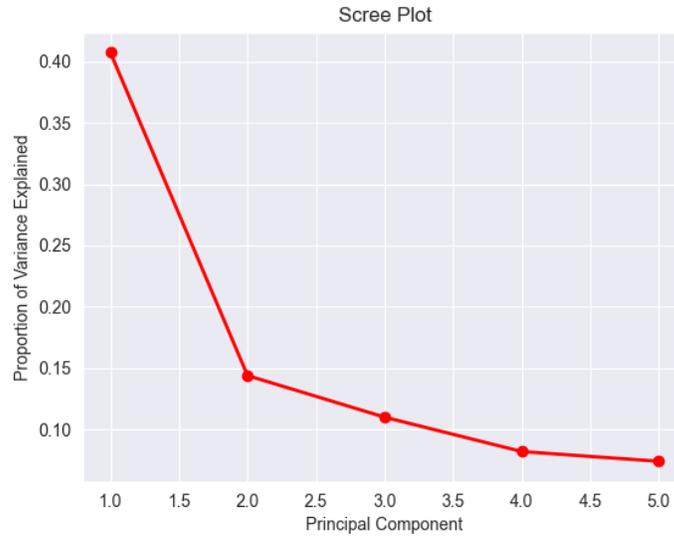


Figure 7. Scree plot of the PCA results using an 11-second time window.

The same modeling analysis was conducted using the four PCs as predictors to test the model performance. Table 5 shows the prediction performance of the five models. The results show the two tree-based models (DT and RF) significantly outperformed the other three linear models, which further indicates the tree-based model can better capture the nonlinear trends in the training and testing set. Between the two tree-based models, the RF model performed significantly better in terms of the training and testing  $R^2$  values. This matches the previous findings that the RF is the best model for predicting takeover performance. However, given that the dimension reduction maintained approximately 70% of the original variability in the training set, the performance of the training and testing sets decreased slightly relative to models without dimension reduction.

Table 5. Prediction model performance of using the 11-second moving window to train the ML models with dimension reduction. All metrics are average values across different randomized splits of the training and testing sets.

	MAE	MSE	RMSE	Training $R^2$	Testing $R^2$
<b>Linear Regression</b>	14.79	301.69	17.37	0.191	0.173
<b>Ridge Regression</b>	13.81	301.82	17.74	0.191	0.172
<b>Lasso Regression</b>	15.56	323.60	17.99	0.135	0.113
<b>Decision Tree</b>	7.26	147.24	12.13	0.817	0.588
<b>Random Forest</b>	<b>5.63</b>	<b>79.12</b>	<b>8.89</b>	<b>0.968</b>	<b>0.783</b>

The PCA result provides implications in identifying the significant groups of variables in predicting takeover performance. To help characterize these groups, the loadings of the top four PCs were extracted in order to identify the most significant factors in predicting takeover performance. The result is presented in Table 6. For the first PC, all significant features are GSR-related indices. Save for a single GSR-related measure, takeover lead time and heart rate-related indices were the primary factors for the second PC. For the third PC, some eye glance measures as well as maximum peak times for GSR and takeover lead time were significant. Lastly, the fourth PC consisted primarily of eye-tracking features and heart rate-related indices. Since the PCs are ordered in decreasing order of percentages of the explained variability, the findings suggest the GSR-related indices were the most important features for identifying driver takeover readiness, followed by heart rate-related indices, takeover lead time, and eye-tracking features.

Table 6. Significant variables with the largest loadings of the top four PCs. Different groups of features are in different font colors to highlight the group's importance.

PC-1	PC-2	PC-3	PC-4
GSR RAW mean	Peak times max	Fixation NDRT	Fixation NDRT duration
GSR RAW max	Heart rate PPG ALG mean	Fixation NDRT duration	Fixation NDRT
GSR RAW min	Heart rate PPG ALG min	Peak times max	Gaze proportion
Peak times max	Heart rate PPG ALG max	Takeover lead time	Fixation road
Peak times std	Takeover lead time		IBI PPG ALG max
GSR conductance CAL max			IBI PPG ALG std
GSR conductance CAL min			
GSR conductance CAL mean			

Note: *blue* = GSR features, *gold* = HR features, *green* = takeover lead time, *red* = eye features.

## Personalized Model for New Drivers

The machine learning model developed in the previous section assumed that one had access to all observations from all participants. However, in practice, when attempting to

predict a new driver's takeover performance, oftentimes there is no prior information or data about the new driver. One viable option is to update a generalized model into a personalized model by continuously incorporating the new driver's data into the machine learning model as they increase their interactions with the vehicle. As such, the personalized model can be fine-tuned to capture individual differences and is expected to achieve better prediction performance.

### *Personalized Prediction Model*

In this step, the proposed model (pre-trained on the training set) was applied to predict the takeover performance of an unobserved driver. For example, if a consumer buys a new Level 3 autonomous vehicle equipped with driver state monitoring features (and the underlying takeover performance prediction model) to monitor their behavior, the initial model will be a generalized model with no prior information about the new driver. In this case, the physiological data from the unobserved driver will be gathered gradually to calibrate the pre-trained generalized model to become a personalized model. If many observations are included in the pre-training, the generalized model can capture more detailed nonlinear trends in the driving population. However, if too much information is used to train the generalized model, when the model is calibrated with new observations it would be very insensitive to the relatively smaller volume of new data. On the other hand, if the generalized model is established with too few observations in the pre-training, the model can be personalized easily with only a small amount of new data but the model performance of the generalized model will be compromised. Thus, there is a tradeoff between the generalization capability and personalization capability. This process was simulated in the current exercise by iteratively adding different percentages of new test data into the training data and evaluating the model performance on the newly collected data. The purpose was to shed light on the design of a personalized model and demonstrate the prediction capability of the proposed personalized model.

The 11-second time window was again employed to explore this concept based on the results from the previous sections. First, the leave-one-out method was used to exclude one participant as the "unobserved" driver. The process was simulated by excluding one participant's data as the testing set. Since each participant only has four takeover events, the step size of the moving window was reduced to 0.25 seconds to generate more testing data. Then, the testing set was divided into ten folds to simulate the process of actively collecting the new driver's data. Therefore, the testing  $R^2$  values were tested on the next fold of the new data to denote they are collected in the next period. Then, the "unobserved" driver's data was gradually included to simulate the online learning process. Specifically, process included the following steps:

1. Train the prediction model on the training set.
2. Collect a new set of data from the new driver.

3. Test on the newly collected set of data.
4. Add the new data to the training set and return to step 1.

Due to the small amount of data observed for each participant, it was unlikely that the new observations would provide sufficient personalization (i.e., there would remain bias towards the generalized model, which includes data from the remaining 31 participants). Therefore, different numbers of training drivers (from 5, 10, 15, and up to 30 drivers) were tested. When there were over 20 drivers' data in the training set, it was observed that the generalized model cannot be adequately personalized to accommodate individual differences after adding new observations, due to the unbalanced training and testing data. On the other hand, if the generalized model was trained on too few drivers' data, it failed to capture the underlying nonlinear trend. That is, it cannot generalize to new drivers' predictions very well. Based on some initial testing, 15 drivers were selected for the training set for this exercise.

### ***Results***

Table 7 presents the result of the model performance of the personalized models built with different percentages of new observations using the 11-second time window and RF model. From Table 7, the testing  $R^2$  values demonstrate the personalized model can be improved as newly observed driver data are gradually added to the generalized model. That is, the adjusted testing  $R^2$  increased as more and more new observations were incorporated into the training set. At the outset, the model can only explain 0.40 of the total variability in the testing set when there are 10% of new observations in the training set. However, the testing significantly improved to 0.59 when another 10% of new observations were added. With personalization, the model can achieve adjusted testing  $R^2$  value of 0.99 when the percentage of new observations gradually increased to 50 percent.

*Table 7. Comparison of model performance of the generalized model and the personalized models built with different percentages of new observations using the 11-second time window and RF model. All metrics are average values across different randomized training and testing sets.*

	MAE	MSE	RMSE	Training $R^2$	Testing $R^2$
<b>Adding 10% of new observations</b>	7.07	135.74	10.49	0.99	0.40
<b>Adding 20% of new observations</b>	0.70	1.52	1.13	0.99	0.59
<b>Adding 30% of new observations</b>	0.59	1.23	0.97	0.99	0.68
<b>Adding 40% of new observations</b>	0.30	0.57	0.70	0.99	0.93
<b>Adding 50% of new observations</b>	0.16	0.22	0.39	0.99	0.99

## Discussion

A human-in-the-loop experiment with 32 participants was conducted, gathering an array of driver physiological data, including heart rate indices, GSR indices, and eye-tracking metrics. A new metric was proposed to evaluate takeover performance based on the Fréchet Distance between theoretically optimal trajectory and real driving trajectory. The FD measures the similarity between two curves and is widely adopted in transportation research (Lyu et al., 2021; Huang et al., 2020). Since the FD calculates the distance of the location and time horizon ordering of the points along the curves, it contains both temporal and spatial characteristics of the takeover performance. Therefore, it is a combined metric that considers the timeliness and quality of the takeover performance.

The aims of this study were to examine the best ML modeling approach and measurement timeframe as well as which measures or indices were most important in characterizing driver takeover readiness. Lastly, general versus personalized models were evaluated.

### *Prediction Models and Measurement Window Length*

In this work, the performance of five ML models was compared in predicting takeover performance. The RF model outperformed all other models when comparing the average testing  $R^2$  values. The moving window length was varied to find the optimal window length for predicting the takeover performance. The results show that the prediction performance on the testing dataset first increases when the window length was increased from 1 second to 11 seconds and then decreases from 12 seconds to 20 seconds, with sharp declines in performance at the 19- and 20-second marks. This suggests that the time window length should be around 11 seconds for the practical application of the proposed ML model (and, in general, the models performed highly in the 9-to-14-second range). Furthermore, the RF and DT models significantly outperformed the other three linear models, demonstrating the capability of capturing the nonlinear trend in the dataset. The outcomes have implications for selecting the range of time windows and modeling approaches to achieve comparable prediction performance.

The wide range of window lengths for near-optimal prediction performance in different settings might be because different physiological signals have different effective window lengths. Some physiological signals (e.g., pupil diameter) perform better with a shorter window size because they change rapidly according to the changes in the driver's cognitive workload (Kramer et al., 2013). Some physiological signals (e.g., heart rate) perform better with a longer window size because it can provide an overall index of the driver's mental state (Solovey et al., 2014). Future research may shed insight into model performance with customized time windows for different physiological signals.

### ***Driver Features and Takeover Readiness***

Thirty-six input features were used to train the prediction model. PCA was then applied to reduce the dimension of the input dimensions to provide insights into the significant groups of variables in the prediction task. That is, measures that were the strongest indicators of driver readiness to takeover control. Using the transformed input features to train the models, the model performance was maintained to an adjusted  $R^2$  value of 0.78 using an 11-second time window. Although the prediction performance did not outperform one using all features, the results show that even a limited number of features can predict 70% of the original variability. This has practical implications since the sensor data is prone to noise and disturbance and the capacity to collect all 36 features is unlikely in real-world applications.

Critically, the most significant variables in the model training were identified from the transformed PCs. Compared with previous studies on using physiological signals to analyze drivers' states and interactions with the driving environment (Mehler et al., 2012; Radlmayr et al., 2014; Wang et al., 2014; Ratwani et al., 2010; Young et al., 2013), the current findings provide implications in the importance of the group of physiological features in indicating driver takeover readiness. Measures derived from GSRs were the strongest predictors of takeover performance in the current study, followed by heart rate indices. As relatively non-intrusive methods for gathering driver physiological measures, it is possible that these might provide valuable inputs or augmentations for future DSM systems. It follows that these results can provide future researchers and practitioners with guidance on how to design driver monitor systems and prioritize the wearable technology and vehicle sensors in a minimally invasive manner to predict drivers' takeover performance in real time.

### ***Personalized Models***

In the last step of this exploration, personalization was shown to significantly improve the model performance without prior information about the new driver. Typically, models have no access to or prior knowledge about a new driver's data. Thus, a generalized prediction model built from the existing drivers' data represents a logical starting point, which is subsequently updated to a personalized version when new observations are available. This process was simulated using personalized models with different percentages of new observations. The results showed that the performance of the personalized models is significantly improved in predicting the new driver's takeover performance when gradually adding about 50% of new observations into the training set. Moreover, it was found that 15 drivers' data can generate a generalized model that ensures both generalization capability to the new driver and the potential of adapting to the personalized models.

### *Key Takeaways*

- A random forest (RF) machine learning approach led to the best model predictions in the current study.
- Pre-takeover measurement windows of 9 to 14 seconds showed the highest model performance, with peak performance for the RF model occurring at 11 seconds.
- Galvanic skin response (GSR) indices were the most important measures in predicting driver performance, followed by heart rate indices.
- Personalized models have great potential for increasing model accuracy in real-world implementations.
- The novel measure of takeover performance, based on the Fréchet Distance (FD), considers both temporal and spatial characteristics and is therefore useful in capturing the timeliness and quality of takeover performance.

## **PART 2: Supporting Driver Attention During Takeover**

---

Part 2 sought to explore avenues to support driver attention during takeover events. It comprises two pieces: (a) identification of potential hazards that co-occur during driver takeover transitions and (b) the design and evaluation of a gaze guidance system to support drivers' noticing of potential hazards during takeovers. The first exercise utilized naturalistic driving data from an existing database, the Integrated Vehicle-Based Safety System (IVBSS) program (Sayer et al., 2011). Ten types of potential hazards were identified using a grounded approach in situations where ADAS systems issued alerts to the driver related to nearby hazards. Researchers documented the co-occurrence of other nearby hazards (ones other than those that triggered the alert) that could conceivably impact driver responses to an alert or takeover situation. The second exercise comprised a driving simulator study aimed at designing and evaluating a gaze guidance system, based on a theoretical model of visual attention (Wickens, 2015). A human-in-the-loop experiment was conducted wherein drivers were exposed to various takeover scenarios with the support of one of two types of gaze guidance system (high and low salience).

### **Identifying Potential Hazards During Takeover Events**

Prior research indicates that failures or degraded performance during takeover transitions are often not because drivers cannot detect the event causing the TOR, but because they have difficulty perceiving the surrounding traffic environment within a short period of time, the so-called "tunnel vision" problem (Zeeb et al., 2016). These nearby objects can become hazards depending on the maneuver chosen by drivers to mitigate the initial hazard event (e.g., swerving to avoid a vehicle ahead but colliding with a vehicle in the blind spot of the adjacent lane). Therefore, the goal of this study was to identify and document the prevalence potential hazards in the surrounding environment that could lead to takeover failures. It is important to note that the focus was not on the potential hazards that directly cause a TOR. Rather, of interest in the project were *other* potential hazards existing in the environment when a TOR is issued. For example, a TOR may be triggered by a slower lead vehicle, and the driver plans to change lanes to avoid a rear-end crash with the lead vehicle. However, at that moment, another vehicle may be located in the adjacent lane. Failing to detect the vehicle in the adjacent lane could lead to a collision.

Due to the absence of a comprehensive naturalistic driving database for conditionally automated vehicles, data from the IVBSS program (Sayer et al., 2011), an existing naturalistic driving study, was utilized. Note that even though the IVBSS program only employed alerts and warnings from lower-level ADAS, such as forward collision warnings, the current exercise used these scenarios as proxies for situations where higher levels of automation are employed (e.g., Level 2 and Level 3 automation).

## Method

The IVBSS program was designed to build and test an integrated in-vehicle crash warning system that includes the forward crash warning, lane departure warning, curve speed warning, and lane change warning (Figure 8; Sayer et al., 2011). Sixteen Honda Accords (ego vehicles) were modified and equipped with seven radar sensors, including one long-range front-facing sensor working at 77 GHz and six sensors working at 24 GHz pointing at adjacent lanes. They were also equipped with a vision system that detected lane boundaries used to determine lane position (LeBlanc et al., 2011). A total of 108 randomly sampled drivers used the test vehicles as a substitute for their personal vehicles over a 40-day period. The IVBSS database for light vehicles represents 213,309 miles, 22,657 trips, and 6,164 hours of driving (Sayer et al., 2011).

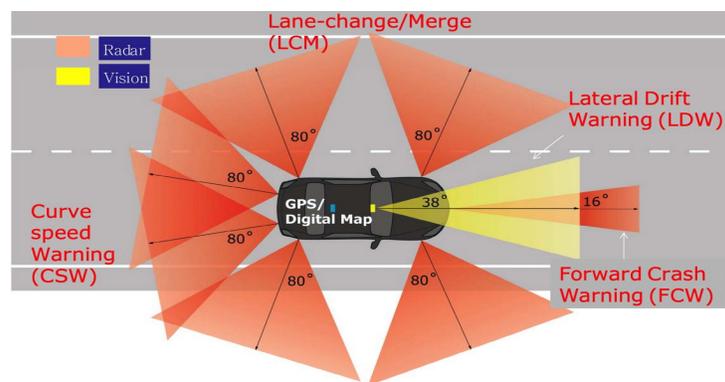


Figure 8. IVBSS Technology (from LeBlanc et al., 2011).

The integrated crash warning system was designed to issue the following warnings (Sayer et al., 2010):

- Forward crash warning (FCW): Warns drivers of the possibility of a rear-end crash with another vehicle.
- Lane-change/merge warning (LCM): Warns drivers of possible crashes with adjacent vehicles, or vehicles approaching in adjacent lanes. LCM includes a blind-spot detection (BSD) function.
- Lateral drift warning (LDW): Warns drivers that they may be drifting out of their lane.
- Curve speed warning (CSW): Warns drivers that they are traveling at an excessive speed and they may not safely negotiate an upcoming curve.

A researcher randomly sampled videos for each type of warning from the IVBSS database. These videos were watched to identify common themes/situations that could be considered potential hazards. These common themes/situations were discussed among the research group to determine if they would be considered potential hazards according to the type of warning. Once a consensus was achieved, a coding rule was developed taking into account the type of warning and the possible maneuvers a driver

might undertake during the warning. Approximately 200 videos were randomly sampled from the IVBSS database for each warning type (duration of each video: 40 seconds). The following selection criteria were also applied: (a) a driver had to have at least 50 miles of driving data, (b) a driver was traveling at a speed of at least 25 mph during the video, and (c) the duration of continuous driving in the video was at least 30 seconds. The query resulted in 194 videos for FCW, 199 videos for LCM, 193 videos for LDW, and 211 videos for CSW, resulting in a total of 797 videos.

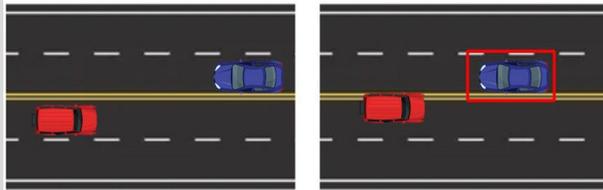
Two researchers coded 240 out of the 797 videos independently and the inter-rater reliability was calculated at 86.5%. After achieving this high inter-rater reliability score, one researcher coded the remaining 557 videos.

**Hazard Scenarios.** This section presents the potential hazards and provides a brief description and a figure depicting the scenario. As noted, the potential hazards of interest are not the ones that cause the warnings (i.e., FCW, LCM, LDW, CSW). Instead, they are other potential hazards in the environment at the time of a warning. The different events and configurations are shown in Table 8 and the coding rules for these events can be found in Appendix A.

Table 8. Hazard scenarios configurations.

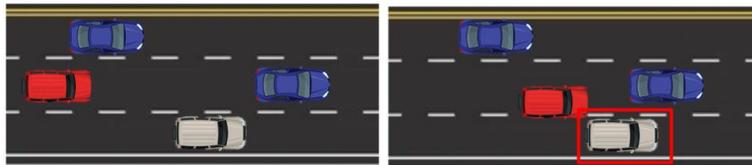
Name	Visualization	Description
Car in Blind Spot (FCW-BS)		<p>A FCW is issued because of the possibility of rear-end crash between the ego vehicle (red) and the leading vehicle (blue). A vehicle in the blind spot (white) is a potential hazard because a possible maneuver of the ego vehicle is to change lanes.</p>
Car in the Adjacent Lane (FCW-AL)		<p>A FCW is issued because of the possibility of rear-end crash between the ego vehicle (red) and the leading vehicle (blue). A car in the adjacent lane (white) within close proximity is a potential hazard because a possible maneuver of the ego vehicle is to change lanes.</p>
Car in Blind Spot After Changing Lanes (LCM-BSCL)		<p>A LCM is triggered as the ego vehicle (red) cuts into an adjacent occupied lane. A car in the blind spot (white) of the ego vehicle after the lane change maneuver is a potential hazard as a possible maneuver of the ego vehicle is to change lanes again.</p>
Car in Adjacent Lane After Changing Lanes (LCM-ALCL)		<p>A LCM is triggered as the ego vehicle (red) cuts into an adjacent occupied lane. After the lane change, a car in the adjacent lane (white) within close proximity to the ego vehicle is a potential hazard because a possible maneuver of the ego vehicle is to change lanes again.</p>
Car in Blind Spot in the Departed Lane (LCM-BSDL)		<p>A LCM is triggered as the ego vehicle (red) cuts into an adjacent occupied lane. After the lane change is completed, a car in the originally departed lane (white) within close proximity to the ego vehicle is a potential hazard because a possible maneuver of the ego vehicle is to change lanes again.</p>

Car in the Opposite Direction (LDW-OD)



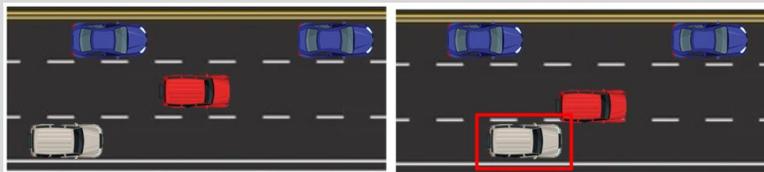
A LDW is triggered as the ego vehicle (red) in the leftmost lane drifts of a two-way road and drifts to the left (i.e. to the lane in the opposite direction). A vehicle traveling in the opposite direction (blue) is a potential hazard having the possibility of causing a head-on collision.

Car in the Adjacent Lane (LDW-AL)



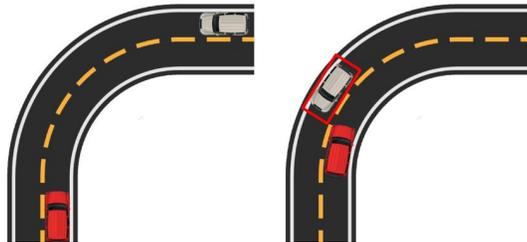
A LDM is triggered by the ego vehicle (red) drifting towards a lane in the same direction. A vehicle traveling in that particular lane (white) within close proximity to the ego vehicle is a potential hazard as a rear-end collision may occur.

Car in Blind Spot (LDW-BS)



A LDW is triggered by the ego vehicle (red) drifting towards a lane in the same direction. A vehicle traveling in that particular lane (white) and is in the blind spot of the ego vehicle is a potential hazard as a rear-end collision may occur.

Car in the Opposite Direction (CSW-OD)



A CSW is triggered because the ego vehicle (red) over speeds when entering a curve. A vehicle traveling in the opposite direction (white) is a potential hazard as the ego vehicle could lose control and go over the center line, causing a head-on or a side collision.

Slower Lead Vehicle (CSW-SLV)



A CSW is triggered because the ego vehicle (red) over speeds when entering a curve. A leading vehicle traveling in the same direction (white) with a slower speed is a potential hazard because the slower vehicle may not be fully visible before entering or at the beginning of the curve, causing a rear-end collision.

## Results and Discussion

Table 9 shows the overall summarized results of the categorization. Across category, the co-occurrence of the alert with another potential hazard ranged from 15.5% to 34.7% (with an overall rate of 24.3%). Co-occurrence was highest for lane change warnings (LCM), followed by forward collision warnings (FCW). In terms of specific situations, LCM-BSCL showed the most occurrences with 14.1% followed by FCW-BS and FCW-AL.

Table 9. Categorization results from video samples.

FCW (N = 194)			LCM (N = 199)			LDW (N = 193)			CSW (N = 211)		
	Count	%									
FCW-BS	25	12.9	LCM-BSCL	28	14.1	LDW-OD	10	5.2	CSW-OD	25	11.8
FCW-AL	24	12.4	LCM-ALCL	20	10.0	LDW-AL	12	6.2	CSW-SVV	21	10.0
			LCM-BSDL	21	10.6	LDW-BS	8	4.1			
<b>Total</b>	<b>59</b>	<b>25.3</b>	<b>Total</b>	<b>69</b>	<b>34.7</b>	<b>Total</b>	<b>30</b>	<b>15.5</b>	<b>Total</b>	<b>46</b>	<b>21.8</b>

Nearly a quarter of the videos sampled in this exercise featured potential secondary hazards, underscoring the importance of considering these elements during takeover transitions. Regarding the specific types of potential hazards, vehicles in the ego vehicle's blind spot were the most frequent situation overall. During the CSW warnings, vehicles in the opposite direction were also of particular concern, as they can lead to head-on collisions, which are the most fatal and represent the highest percentage of total fatal crashes (Insurance Information Institute, 2019).

## Design and Evaluation of a Gaze Guidance System

The above exercise served to quantify and characterize the types of potential hazards that may exist in the surrounding traffic environment during takeover transitions. In this study, a gaze guidance system was designed and evaluated aiming to support the noticing of potential hazards. The gaze guidance system was motivated by the N-SEEV (Salience, Effort, Expectancy, Value) model of selective visual attention, focusing on the salience factor (Wickens, 2015). That is, the system was intended to cue or highlight important, safety-critical information for drivers in order to guide their attention to this information during their resumption of control from automation. Different levels of system salience were evaluated. Based on the potential hazards identified in previous section and considerations regarding the implementation in a simulator study, the FCW-BS scenario was chosen for the current study.

## **Method**

**Participants.** A total of  $N = 12$  participants (average age = 25.8 years,  $SD = 4.4$  years, 6 females, 6 males) with normal or corrected-to-normal vision participated in the experiment. Each participant received a payment of \$30 for their participation. Participants had not participated in the study in Part 1.

**Apparatus and Stimuli.** The same driving simulator and the same NDRT task as described in Part 1 were used for the current experiment. During the experiment, participants wore the Pupil Core Glasses eye-tracking system (Pupil Labs, Germany) that provided real-time gaze and pupil data at a sampling rate of 200 Hz. The vehicle was programmed to simulate the behavior of an SAE Level 3 automation, which handled the longitudinal and lateral control and navigation, and responded to traffic elements. Participants could press the button on the steering wheel to activate the automated mode, which was indicated by an auditory sound and a blue light on the dashboard. Whenever the AV reached its system limit, a take-over request, consisting of a spoken auditory warning (“Takeover”) and the disappearance of the blue light on the dashboard, was issued.

**Gaze Guidance System.** Two versions of the gaze guidance system were evaluated, which varied by the salience of the cue (high salience or low salience). For both high and low salience conditions, the side mirror was highlighted with a red bounding box when a potential hazard (i.e., the blue vehicle) was located near the ego vehicle and was about to enter the blind spot of the ego vehicle at the moment of TOR. The potential hazard was programmed to drive at a fixed speed. Once the driver slowed down after hearing the TOR, the potential hazard vehicle would enter the blind spot of the ego vehicle and eventually bypass the ego vehicle (if there was no crash). In the high salience condition, the red box would flash 20 times over 4 seconds (5 Hz), whereas in the low salience condition, the red bounding box would appear for 4 seconds without flashing.



*Figure 9. Illustration of gaze guidance system: control condition (left), low salience guidance (right).*

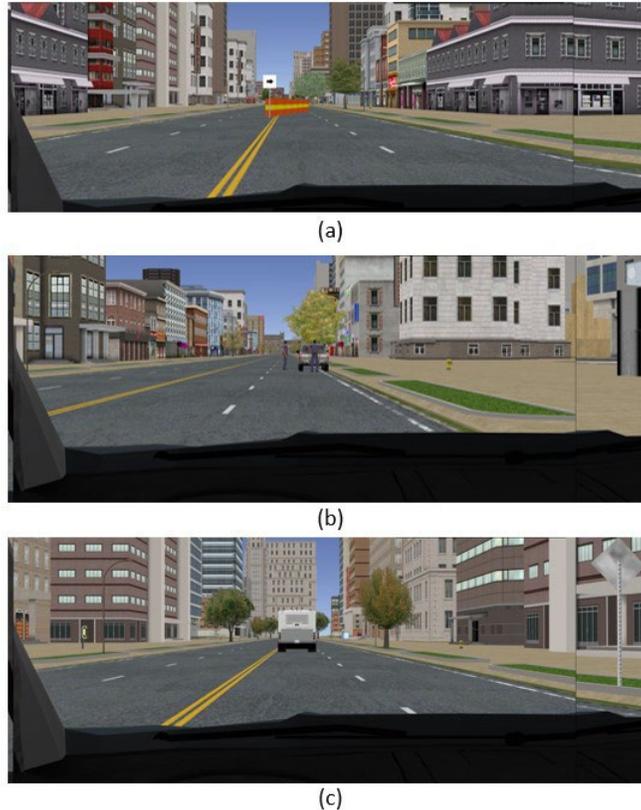
**Procedure.** Upon arrival, participants provided informed consent and filled out a demographic survey. The participants were then given an introduction that describes the content of the experiment. Next, participants received a training session to get familiar with the driving simulator and the Tetris game. During the training, participants

practiced how to drive, change lanes, and engage/disengage in the automated driving mode. They were also presented with visual and auditory alerts for takeover requests. Participants were then asked to drive the simulator until they felt comfortable handling the simulator controls. After the training, participants were fitted with the eye tracker.

Each participant experienced all three salience conditions (high salience, low salience, and control) and encountered three takeover events in the experiment. The three events are illustrated in Table 10 and Figure 10. The sequence of the three conditions was counterbalanced using a balanced Latin square design.

*Table 10. Descriptions of takeover events.*

<b>Event</b>	<b>Scenario Type</b>	<b>Description</b>
1	Construction Ahead	Urban two-lane road—construction blocking the lane (traffic cones and equipment) requiring that driver change lanes to the adjacent (vacant) lane.
2	Police Officer	Two-lane road—police officer is inspecting a vehicle partially obstructing the lane, requiring driver to change lanes.
3	Stationary Bus Ahead	Urban two-lane road—a bus is stopped unexpectedly in the lane ahead, requiring the driver to change lanes.



*Figure 10. Illustration of takeover events: (a) construction ahead, (b) police vehicle on shoulder, (c) stationary bus ahead*

Each driving block began with the command to activate the automated driving mode was given. After that, there was an NDRT phase where participants were asked to play the Tetris game. Participants were informed that there was no need to monitor the environment when the AV was in automated driving mode. Once a TOR was issued, participants were required to take over control of the vehicle immediately. Participants were instructed to comply with all the traffic laws when they drove manually. They were informed that the speed limit was 35 mph. Participants could hand back the control to the AV after they negotiated the driving situation. When the AV was re-engaged, participants completed a brief subjective questionnaire verbally. At the end of the experiment, participants ranked their preferences for the three designs.

**Dependent Measures.** Several dependent variables were measured including the occurrence of crash/near crash, noticing time of the potential hazard, and subjective measurements. Given the specific design of the experiment, a rear-end crash/near crash occurred when the ego vehicle tried to change lanes without noticing the hazard vehicle. There was no crash when the ego vehicle waited for the target vehicle to pass and afterward switched lanes. The noticing time of the potential hazard vehicle is the time taken for the driver to notice the hazard vehicle. It was calculated as the time between the issue of TOR and the first fixation on the side mirror (Figure 11).

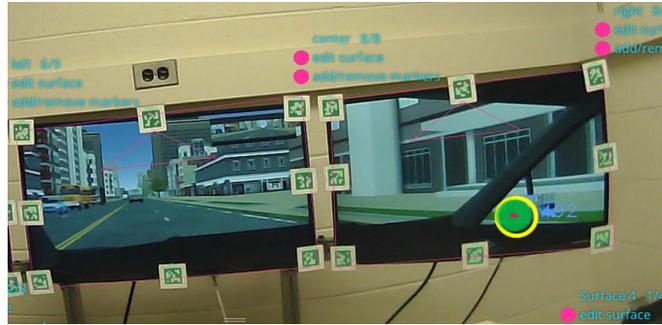


Figure 11. Illustration of the first fixation on the side mirror.

Additionally, participants' perceived takeover readiness, their level of engagement in the NDRT before TOR, and their perceived usefulness and ease of use for the gaze guidance system were measured (Table 11). At the end of the experiment, participants were asked to rank their preference for the three designs (i.e., high salience, low salience, and control).

Table 11. Subjective measures of perceived readiness, level of engagement, perceived usefulness, and ease of use for the gaze guidance system.

Measure	Questions or Items
Perceived readiness	How was your takeover performance?
Level of engagement	I was totally absorbed by the Tetris task before the takeover request.
Perceived usefulness	The gaze guidance system helped me recognize the situation around me. Using the gaze guidance system will improve my takeover performance. Using the gaze guidance system increases my safety during takeover transitions.
	Using the gaze guidance system enhances my effectiveness to take over control of the vehicle.
	I find the gaze guidance system to be useful during takeover transitions.
Perceived ease of use	The gaze guidance system is clear and understandable.
	Interacting with the gaze guidance system does not require a lot of mental effort.
	I find the gaze guidance system easy to use.

## Results

The experiment used a one-way within-subjects design with the salience level (high salience, low salience, and control) as the independent variable. Table 12 shows the occurrences of crashes/near-crashes and the noticing time as participants encountered the three events. Despite the use of a Latin square design, there were significant learning effects as participants experienced more takeover events. Therefore, the current analysis focuses on the first event.

**Crashes and Near Crashes.** A  $\chi^2$  test compared the likelihood of crashes/near crashes occurrence between the three salience conditions. The result showed a significant difference,  $\chi^2(2) = 6.0$ ,  $p = .05$ ; there was a lower likelihood of crashes/near crashes with the high salience design, compared to the low salience design and the control condition.

Table 12. Occurrence of crash/near crash and noticing time across participants and event.

Participant #	Event 1			Event 2			Event 3		
	Salience	Crash	Noticing Time (sec)	Salience	Crash	Noticing Time (sec)	Salience	Crash	Noticing Time (sec)
1	High	0	4.04	Control	0	2.94	Low	0	1.34
2	Low	1	-	High	0	1.31	Control	0	1.18
3	Control	1	1.83	Low	0	1.59	High	0	1.23
4	Low	0	4.39	Control	0	4.03	High	0	2.73
5	Control	1	1.80	High	0	1.87	Low	0	0.90
6	High	0	0.84	Control	0	1.07	Low	0	0.89
7	Low	NC	-	High	1	1.86	Control	0	2.21
8	Control	1	1.99	Low	0	1.46	High	0	1.07
9	Low	1	-	Control	0	1.10	High	0	0.66
10	High	0	4.56	Low	0	2.24	Control	0	2.35
11	High	0	1.14	Low	0	1.35	Control	0	1.66
12	Control	0	3.01	High	0	1.76	Low	0	1.03

Note: Crash: 1=Yes, 0=No, NC=Near Crash.

**Noticing of Potential Hazard.** A one-way repeated-measures analysis of variance (ANOVA) was used to compare the noticing time between the three conditions, which showed a non-significant difference ( $F(2, 6) = .98, p = .43$ ).

**Subjective Ratings.** Table 13 tabulates the mean and standard deviation (SD) of perceived readiness, engagement in NDRT, perceived usefulness, and perceived ease of use. Although the high salience condition showed nominally higher ratings of perceived readiness, usefulness, and ease of use, the one-way ANOVAs revealed non-significant differences (perceived readiness ( $F(2, 9) = 2.29, p = .16$ ); engagement in NDRT ( $F(2, 9) = .00, p = .81$ ); perceived usefulness ( $F(1, 6) = 1.34, p = .29$ ); and perceived ease of use ( $F(1, 6) = 2.23, p = .12$ )).

Table 13. Mean and SD of perceived readiness, engagement in NDRT, perceived usefulness, and perceived ease of use.

	Perceived Performance	Engagement in NDRT	Perceived usefulness	Perceived ease of use
High salience	5.8 (1.9)	5.5 (2.4)	4.6 (1.2)	5.0 (1.2)
Low salience	3.0 (1.4)	6.0 (2.0)	3.5 (1.7)	3.1 (1.8)
Control	3.5 (2.4)	6.0 (1.6)		

Figure 12 shows the ranking of the three salience conditions (1st – most preferred, 3rd – least preferred). a Friedman’s test shows a significant difference between the ranks ( $\chi^2(2) = 18.5, p < .001$ ). The high salience design was preferred the most, followed by the low salience design and the control condition.

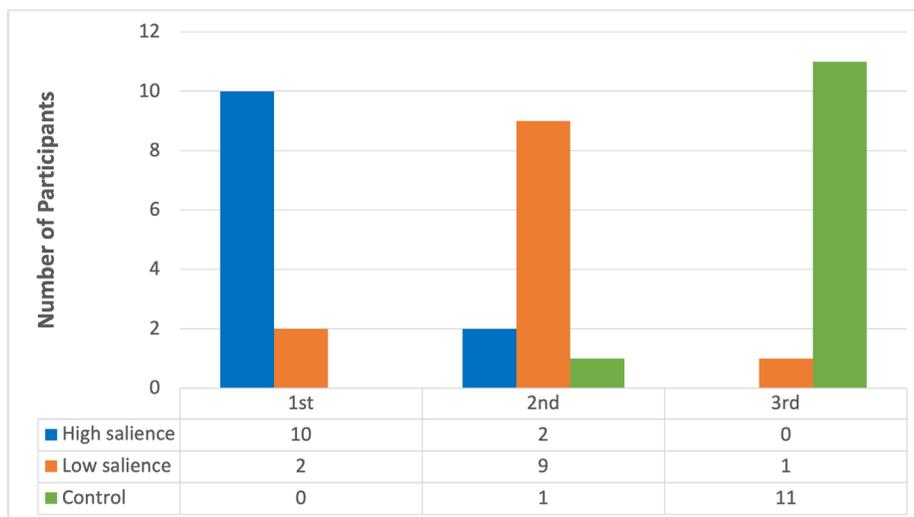


Figure 12. Rankings of preferred system types.

## *Discussion*

The purpose of this study was to showcase the potential of a gaze guidance system in improving takeover readiness and performance. Despite the limited number of participants, the study yielded noteworthy findings. The high salience design significantly reduced the likelihood of crashes, with no incidents reported among participants in this group. In contrast, three crashes were observed in both the low salience and control conditions. The current analysis did not reveal a difference between the low salience and control conditions. Notably, three out of four participants in the low salience group failed to check the side mirror, suggesting that this version of the system was not sufficiently salient. This underscores the potential benefits of an attention-grabbing mechanism to redirect the driver's attention to the potential hazards. Furthermore, the participants' evaluation indicated a trend in favor of the high salience condition, with higher perceived performance, usefulness, and ease of use compared to the low salience condition.

Although this study was preliminary, this proof-of-concept study lays the foundation for future research and development in this area of driver support during takeover events. Attentional cueing and guidance is an effective means of highlighting task-critical information in the world. Further studies with larger sample sizes under a wider range of conditions are needed to further validate the current results.

## General Discussion

---

Vehicle automation continues to evolve and change the nature of the driver's role, from an active operator to more of a system supervisor. However, when drivers are decoupled from the operational control of the vehicle, they often have difficulty taking back control—especially in situations that the automation is not able to handle (Zhou et al., 2020; Petersen et al., 2019; Molnar et al., 2017). Thus, understanding how to monitor and support drivers during such takeover situations is imperative. The current set of studies aimed to provide insight into indices of driver takeover readiness as well as attentional support in executing a takeover. A number of high-level takeaways are highlighted and/or re-iterated:

- In examining driver physiological data in real-time, a random forest (RF) machine learning approach led to the best model predictions. Driver state monitoring systems (and their underlying algorithms) should consider such approaches.
- Pre-takeover measurement windows of 9 to 14 seconds showed the highest model performance, with peak performance for the RF model occurring at 11 seconds. DSM systems should strive to incorporate and/or validate their own outcomes using such time frames.
- Galvanic skin response (GSR) indices were the most important measures in predicting driver performance, followed by heart rate indices. As indicators of driver takeover readiness, these measures should be considered for inclusion in DSM systems as part of a suite of measures.
- Personalized models have great potential for increasing model accuracy in real-world implementations. Generalized models that learn from new users have the potential to increase their accuracy and real-world utility.
- The novel measure of takeover performance, based on the Fréchet Distance, considers both temporal and spatial characteristics and is therefore useful in capturing the timeliness and quality of takeover performance.
- Gaze guidance (attentional) support is one viable approach to helping drivers during takeover events. More work and innovation in this area is merited.

## References

---

- Abbas, M. A., Milman, R., & Eklund, J. M. (2017). Obstacle avoidance in real time with nonlinear model predictive control of autonomous vehicles. *Canadian Journal of Electrical and Computer Engineering*, 40(1), 12–22.  
<https://doi.org/10.1109/CJECE.2016.2609803>
- Alt, H. & Godau, M. (1995). Computing the Fréchet distance between two polygonal curves. *International Journal of Computational Geometry & Applications*, 5(01n02), 75–91. <https://doi.org/10.1142/S0218195995000064>
- Andersson, J. A. E., Gillis, J., Horn, G., Rawlings, J. B., & Diehl, M. (2019). CasADi – A software framework for nonlinear optimization and optimal control. *Mathematical Programming Computation*, 11(1), 1–36.  
<https://doi.org/10.1007/s12532-018-0139-4>
- Benedek, M. & Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods*, 190(1), 80–91.  
<https://doi.org/10.1016/j.jneumeth.2010.04.028>
- Braunagel, C., Rosenstiel, W., & Kasneci, E. (2017). Ready for take-over? A new driver assistance system for an automated classification of driver take-over readiness. *IEEE Intelligent Transportation Systems Magazine*, 9(4), 10–22.  
<https://doi.org/10.1109/MITS.2017.2743165>
- Clark, H. & Feng, J. (2017). Age differences in the takeover of vehicle control and engagement in non-driving-related activities in simulated driving with conditional automation. *Accident Analysis & Prevention*, 106, 468–479.  
<https://doi.org/10.1016/j.aap.2016.08.027>
- Du, N., Yang, X. J., & Zhou, F. (2020a). Psychophysiological responses to takeover requests in conditionally automated driving. *Accident Analysis & Prevention*, 148, 105804.  
<https://doi.org/10.1016/j.aap.2020.105804>
- Du, N., Zhou, F., Pulver, E., Tilbury, D. M., Robert, L. P., Pradhan, A. K., & Yang, X. J. (2020b). Examining the effects of emotional valence and arousal on takeover performance in conditionally automated driving. *Transportation Research Part C: Emerging Technologies*, 112, 78–87. <https://doi.org/10.2139/ssrn.3518015>
- Du, N., Zhou, F., Pulver, E. M., Tilbury, D. M., Robert, L. P., Pradhan, A. K., & Yang, X. J. (2020c). Predicting driver takeover performance in conditionally automated driving. *Accident Analysis & Prevention*, 148, 105748.  
<https://doi.org/10.1016/j.aap.2020.105748>
- Eiter, T. & Mannila, H. (1994). *Computing Discrete Fréchet Distance* (Technical Report CD-TR 94/64). Vienna, Austria: Technische Universität Wien.

- Eriksson, A., Petermeijer, S. M., Zimmermann, M., De Winter, J. C., Bengler, K. J., & Stanton, N. A. (2018). Rolling out the red (and green) carpet: Supporting driver decision making in automation-to-manual transitions. *IEEE Transactions on Human-Machine Systems*, 49(1), 20–31. <https://doi.org/10.1109/THMS.2018.2883862>
- Eriksson, A. & Stanton, N. A. (2017). Takeover time in highly automated vehicles: noncritical transitions to and from manual control. *Human Factors*, 59(4), 689–705. <https://doi.org/10.1177/0018720816685832>
- Febbo, H., Liu, J., Jayakumar, P., Stein, J. L., & Ersal, T. (2017). Moving obstacle avoidance for large, high-speed autonomous ground vehicles. *2017 American Control Conference (ACC)* (pp. 5568–5573). IEEE. <https://doi.org/10.23919/ACC.2017.7963821>
- Gold, C., Happee, R., & Bengler, K. (2018). Modeling take-over performance in level 3 conditionally automated vehicles. *Accident Analysis & Prevention*, 116, 3–13. <https://doi.org/10.1016/j.aap.2017.11.009>
- Gold, C., Körber, M., Lechner, D., & Bengler, K. (2016). Taking over control from highly automated vehicles in complex traffic situations: the role of traffic density. *Human Factors*, 58(4), 642–652. <https://doi.org/10.1177/0018720816634226>
- Happee, R., Gold, C., Radlmayr, J., Hergeth, S., & Bengler, K. (2017). Take-over performance in evasive manoeuvres. *Accident Analysis & Prevention*, 106, 211–222. <https://doi.org/10.1016/j.aap.2017.04.017>
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., ..., & Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585(7825), 357–362. <https://doi.org/10.1038/s41586-020-2649-2>
- Helldin, T., Falkman, G., Riveiro, M., & Davidsson, S. (2013). Presenting system uncertainty in automotive UIs for supporting trust calibration in autonomous driving. *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '13* (pp. 210–217), New York: ACM. <https://doi.org/10.1145/2516540.2516554>
- Huang, Z., Wang, P., & Liu, Y. (2020). Statistical characteristics and transportation mode identification of individual trajectories. *International Journal of Modern Physics B*, 34(10), 2050092. <https://doi.org/10.1142/S0217979220500927>
- Insurance Information Institute (2019). *Facts + Statistics: Highway Safety*. Available at: <https://www.iii.org/fact-statistic/facts-statistics-highway-safety>.
- Kramer, S. E., Lorens, A., Coninx, F., Zekveld, A. A., Piotrowska, A., & Skarzynski, H. (2013). Processing load during listening: The influence of task characteristics on the pupil response. *Language and Cognitive Processes*, 28(4), 426–442. <https://doi.org/10.1080/01690965.2011.642267>

- Lamble, D., Laakso, M., & Summala, H. (1999). Detection thresholds in car following situations and peripheral vision: Implications for positioning of visually demanding in-car displays. *Ergonomics*, 42(6), 807–815. <https://doi.org/10.1080/001401399185306>
- LeBlanc, D., Sayer, J. R., Bao, S., Bogard, S., Buonarosa, M. L., Blankespoor, A., & Funkhouser, D. (2011). Driver acceptance and behavioral changes with an integrated warning system: Key finding from the IVBSS FOT. *22nd International Technical Conference on the Enhanced Safety of Vehicles (ESV)* (pp. 1–10).
- Li, S., Blythe, P., Guo, W., and Namdeo, A. (2018). Investigation of older driver's takeover performance in highly automated vehicles in adverse weather conditions. *IET Intelligent Transport Systems*, 12(9), 1157–1165. <https://doi.org/10.1049/iet-its.2018.0104>
- Lyu, C., Wu, X., Liu, Y., & Liu, Z. (2021). A partial-Fréchet-distance-based framework for bus route identification. *IEEE Transactions on Intelligent Transportation Systems*, 23(7), 9275–9280. <https://doi.org/10.1109/TITS.2021.3069630>
- McKinney, W. (2010). Data structures for statistical computing in python. *Proceedings of the 9th Python in Science Conference*, 445, 51–56. <https://doi.org/10.25080/Majora-92bf1922-00a>
- Mehler, B., Reimer, B., & Coughlin, J. F. (2012). Sensitivity of physiological measures for detecting systematic variations in cognitive demand from a working memory task: An on-road study across three age groups. *Human Factors*, 54(3), 396–412. <https://doi.org/10.1177/0018720812442086>
- Meng, X., Fu, H., Liu, G., Zhang, L., Yu, Y., Hu, W., & Cheng, E. (2019). Multi-feature fusion: a driver-car matching model based on curve comparison. *IEEE Access*, 7, 83526–83535. <https://doi.org/10.1109/ACCESS.2019.2923795>
- Merat, N., Jamson, A. H., Lai, F. C., & Carsten, O. (2012). Highly automated driving, secondary task performance, and driver state. *Human Factors*, 54(5), 762–771. <https://doi.org/10.1177/0018720812442087>
- Molnar, L. J., Pradhan, A. K., Eby, D. W., Ryan, L. H., Louis, R. M. S., Zakrajsek, J., Ross, B., Lin, B. T., Liang, C., Zalewski, B., and Zhang, L. (2017). *Age-Related Differences in Driver Behavior Associated with Automated Vehicles and the Transfer of Control Between Automated and Manual Control: A Simulator Evaluation* (Technical Report UMTRI2017-4). Ann Arbor, MI: University of Michigan Transportation Research Institute.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ..., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.

- Petersen, L., Robert, L., Yang, X. J., and Tilbury, D. (2019). Situational awareness, driver's trust in automated driving systems and secondary task performance. *SAE International Journal of Connected and Automated Vehicles*, 2(2). <https://doi.org/10.2139/ssrn.3345543>
- Radlmayr, J., Gold, C., Lorenz, L., Farid, M., & Bengler, K. (2014). How traffic situations and non-driving related tasks affect the take-over quality in highly automated driving. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 58, 2063–2067. <https://doi.org/10.1177/1541931214581434>
- Ratwani, R. M., McCurry, J. M., & Trafton, J. G. (2010). Single operator, multiple robots: an eye movement based theoretic model of operator situation awareness. *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 235–242). IEEE. <https://doi.org/10.1109/HRI.2010.5453191>
- Sayer, J. R., Buonarosa, M. L., Bao, S., Bogard, S. E., LeBlanc, D. J., Blankespoor, A. D., Funkhouser, D. S., & Winkler, C. B. (2010). *Integrated Vehicle-Based Safety Systems Light-Vehicle Field Operational Test Methodology and Results Report* (Technical Report UMTRI-2010-30): Ann Arbor, MI: University of Michigan, Transportation Research Institute.
- Sayer, J. R., LeBlanc, D. J., Bogard, S., Funkhouser, D., Bao, S., Buonarosa, M. L., & Blankespoor, A. (2011). *Integrated Vehicle-Based Safety Systems Field Operational Test Final Program Report* (Technical Report DOT HS 811 482). Washington, D.C.: U.S. Department of Transportation.
- Solovey, E. T., Zec, M., Garcia Perez, E. A., Reimer, B., & Mehler, B. (2014). Classifying driver workload using physiological and driving performance data: Two field studies. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 4057–4066). New York: ACM. <https://doi.org/10.1145/2556288.2557068>
- Wächter, A. & Biegler, L. T. (2006). On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical Programming*, 106, 25–57. <https://doi.org/10.1007/s10107-004-0559-y>
- Wan, J. & Wu, C. (2018). The effects of lead time of take-over request and nondriving tasks on taking-over control of automated vehicles. *IEEE Transactions on Human-Machine Systems*, 99, 1–10. <https://doi.org/10.1109/THMS.2018.2844251>
- Wang, Y., Reimer, B., Dobres, J., & Mehler, B. (2014). The sensitivity of different methodologies for characterizing drivers' gaze concentration under increased cognitive demand. *Transportation Research Part F: Traffic Psychology and Behaviour*, 26, 227–237. <https://doi.org/10.1016/j.trf.2014.08.003>
- Wickens, C. D. (2015). Noticing events in the visual workplace: The SEEV and NSEEV models. *Cambridge Handbooks in Psychology* (pp. 749-768). Cambridge University Press. <https://doi.org/10.1017/CBO9780511973017.046>

- Wintersberger, P., Riener, A., Schartmüller, C., Frison, A.-K., & Weigl, K. (2018). Let me finish before I take over: Towards attention aware device integration in highly automated vehicles. *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 53–65). New York: ACM. <https://doi.org/10.1145/3239060.3239085>
- Young, K. L., Salmon, P. M., and Cornelissen, M. (2013). Missing links? The effects of distraction on driver situation awareness. *Safety Science*, *56*, 36–43. <https://doi.org/10.1016/j.ssci.2012.11.004>
- Zeeb, K., Buchner, A., & Schrauf, M. (2016). Is take-over time all that matters? The impact of visual-cognitive load on driver take-over quality after conditionally automated driving. *Accident Analysis & Prevention*, *92*, 230–239. <https://doi.org/10.1016/j.aap.2016.04.002>
- Zeeb, K., Härtel, M., Buchner, A., & Schrauf, M. (2017). Why is steering not the same as braking? The impact of non-driving related tasks on lateral and longitudinal driver interventions during conditionally automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *50*, 65–79. <https://doi.org/10.1016/j.trf.2017.07.008>
- Zhou, F., Yang, X. J., and Zhang, X. (2020). Takeover transition in autonomous vehicles: A YouTube study. *International Journal of Human-Computer Interaction*, *36*(3), 295–306. <https://doi.org/10.1080/10447318.2019.1634317>

## Appendix A: Coding Rules for Hazard Alert Conditions

Type	Coding Rules
Car in Blind Spot (FCW-BS)	<ul style="list-style-type: none"> <li>The ego vehicle is approaching a lead vehicle and triggers an FCW</li> <li>There is a car in the blind spot that triggers the blind spot signal during the time the alarm message is displayed before the ego vehicle changes lanes</li> <li>Classify it as FCW-BS</li> </ul>
Car in the Adjacent Lane (FCW-AL)	<ul style="list-style-type: none"> <li>There is a vehicle in the adjacent lane at the moment of a FCW</li> <li>The vehicle is within close proximity to the ego vehicle, in between the leading vehicle causing the FCW and the ego vehicle</li> <li>Classify it as FCW-AL</li> </ul>
Car in Blind Spot After Changing Lanes (LCM-BSCL)	<ul style="list-style-type: none"> <li>LCM is triggered when the ego vehicle cuts into the adjacent lane</li> <li>The blind spot signal (left blind spot if the ego vehicle switches to the left lane, right blind spot if the ego vehicle switches to the right lane) is activated due to the vehicle in the blind spot</li> <li>Classify it as LCM-BSCL</li> </ul>
Car in Adjacent Lane After Changing Lanes (LCM-ALCL)	<ul style="list-style-type: none"> <li>An LCM is triggered when the ego vehicle cuts into the next lane</li> <li>After the lane changing maneuver is completed, there is a car in the adjacent lane (left lane if the ego vehicle switched to the left lane, right lane if the ego vehicle switched to the right lane)</li> <li>Rear portion of the car overlaps with the ego vehicle</li> <li>Classify it as LCM-ALCL</li> </ul> <p>OR</p> <ul style="list-style-type: none"> <li>An LCM is triggered when the ego vehicle cuts into the next lane</li> <li>After the lane changing maneuver is completed, there is a car in the adjacent lane (left lane if the ego vehicle switched to the left lane, right lane if the ego vehicle switched to the right lane)</li> <li>Apply a 2 seconds rule from the time the hazard signal appears; the car is within 2 seconds and is traveling slower than the ego vehicle.</li> <li>Classify it as LCM-ALCL</li> </ul>
Car in Blind Spot on the Departed Lane (LCM-BSDL)	<ul style="list-style-type: none"> <li>An LCM is triggered and the ego vehicle cuts into the next lane</li> <li>The blind spot signal is activated by the car in the previous lane</li> <li>Classify it as LCM-BSDL</li> </ul>
Car in the Opposite Direction (LDW-OD)	<ul style="list-style-type: none"> <li>An LDM is triggered by the ego vehicle drifting towards the boundary between ego vehicle's lane and those in the opposite direction (e.g., double yellow lines)</li> <li>There is a car traveling in the opposite direction on the lane the ego vehicle is drifting toward during LDM</li> <li>Classify it as LDW-OD</li> </ul>

---

<p>Car in the Adjacent Lane (LDW-AL)</p>	<ul style="list-style-type: none"> <li>• An LDW is triggered by the ego vehicle drifting to an adjacent lane in the same direction</li> <li>• There is a car in the adjacent lane (left if the ego vehicle drifts to the left, right if the ego vehicle drifts to the right)</li> <li>• Any portion of the car overlaps with the ego vehicle</li> <li>• Classify it as LDW-AL</li> </ul> <p>OR</p> <ul style="list-style-type: none"> <li>• An LDW is triggered by the ego vehicle drifting to an adjacent lane in the same direction</li> <li>• There is a car in the adjacent lane (left if the ego vehicle drifts to the left, right if the ego vehicle drifts to the right)</li> <li>• Apply a 2 seconds rule from the time the drift signal appears; the car is within 2 seconds and is traveling slower than ego vehicle</li> <li>• Classify it as LDW-AL</li> </ul>
<p>Car in Blind Spot (LDW-BS)</p>	<ul style="list-style-type: none"> <li>• An LDW is triggered by the ego vehicle drifting to an adjacent lane in the same direction</li> <li>• A car in the lane to which the ego vehicle is drifting triggers the blind spot signal</li> <li>• Classify it as LDW-BS</li> </ul>
<p>Car in the Opposite Direction (CSW-OD)</p>	<ul style="list-style-type: none"> <li>• The ego vehicle is traveling fast while approaching/handling a curve and triggers the alarm</li> <li>• There is a car traveling in the opposite direction when the alert is visible</li> <li>• Classify it as CSW-OD</li> </ul>
<p>Slower Lead Vehicle (CSW-SLV)</p>	<ul style="list-style-type: none"> <li>• A CSW is triggered when the ego vehicle enters a curved road</li> <li>• The ego vehicle does not brake and there is a lead vehicle on the curve road traveling slower than the ego vehicle</li> <li>• Classify it as CSW-SLV</li> </ul> <p>OR</p> <ul style="list-style-type: none"> <li>• A CSW is triggered when the ego vehicle enters a curved road</li> <li>• The ego vehicle brakes during the alarm and the lead vehicle in the curve is traveling slower than the ego vehicle during the 3 seconds before the brake.</li> <li>• Classify it as CSW-SLV</li> </ul>

---