Building Contextualized Trust Profiles in Conditionally Automated Driving

Lilit Avetisyan, Jackie Ayoub, X. Jessie Yang, *Associate Member, IEEE*, and Feng Zhou, *Senior Member, IEEE*

***Abstract*—Trust is crucial for ensuring the safety, security, and widespread adoption of automated vehicles (AVs), and if trust is lacking, drivers and the general public may hesitate to embrace this technology. This research seeks to investigate contextualized trust profiles in order to create personalized experiences for drivers in AVs with varying levels of reliability. A driving simulator experiment involving 70 participants revealed three distinct contextualized trust profiles (i.e., *confident copilots*, *myopic pragmatists*, and *reluctant automators*) identified through K-means clustering, and analyzed in relation to drivers' dynamic trust, dispositional trust, initial learned trust, personality traits, and emotions. The experiment encompassed eight scenarios where participants were requested to take over control from the AV in three conditions: a control condition, a false alarm condition, and a miss condition. To validate the models, a multinomial logistic regression model was constructed using the shapley additive explanations explainer to determine the most influential features in predicting contextualized trust profiles, achieving an F1-score of 0.90 and an accuracy of 0.89. In addition, an examination of how individual factors impact contextualized trust profiles provided valuable insights into trust dynamics from a user-centric perspective. The outcomes of this research hold significant implications for the development of personalized in-vehicle trust monitoring and calibration systems to modulate drivers' trust levels, thereby enhancing safety and user experience in automated driving.**

***Index Terms*—Automated vehicles (AVs), emotion, personality traits, contextualized trust profiles.**

I. Introduction

AUTOMATED vehicles (AVs) possess the potential to revolutionize the transportation sector by offering safer and more efficient modes of transportation [1]. Nevertheless, the widespread acceptance and implementation of AVs heavily rely on users' trust in this technology. In the absence of trust, drivers may exhibit reluctance toward adopting AVs, thereby impeding or even hindering the deployment of this transformative technology. Consequently, it becomes imperative to design AVs in a manner that tailoring AV experiences for individual users (or user models) based on their trust levels in AVs, encouraging users to accept and embrace them. Moreover, the ability to perform non-driving related tasks (NDRTs) is one of the promises of AVs, allowing drivers to act like passengers. Conversely, lack of trust may lead to hesitation in fully relying on automation, resulting in divided attention between driving and NDRTs. While trust enables NDRTs, drivers must remain vigilant to take over control when needed. Therefore, balancing trust and vigilance is critical [2], [3].

Thus, many researchers have conducted studies on trust in AVs and other types of automation, with a particular focus on factors that contribute to and influence trust, such as system transparency, reliability, and performance [4], [5]. For example, both Ayoub et al. [4] and Azevedo-Sa et al. [6] demonstrated that when an AV exhibited a high level of system reliability participants were more inclined to trust the vehicle. In addition, several studies (e.g., [5], [7], and [8]) highlighted the importance of enhancing system transparency. By providing information about the system's status and explanations for automation failures, trust in automation and AVs was increased.

The aforementioned studies have provided valuable insights to aid in the design of automated systems. However, what appears to be lacking and inconsistent is the consideration of the impact of individual differences on trust in AVs, including factors such as age, driving experience, knowledge of automation systems, personality traits, and emotions (e.g., [9], [10], and [11]). For instance, empirical studies (e.g., [12]) indicated that older adults exhibit a higher tendency to overtrust automated systems compared to younger age groups. Conversely, survey findings revealed that older participants (60 years and above: 45.2%) expressed greater concerns, and therefore, exhibited lower levels of trust in AVs compared to their younger counterparts (18–29 years old: 26.1%) [13]. It is important to note that individuals of ages often employ distinct strategies influenced by various factors such as knowledge, personality traits, and driving experience. Moreover, the specific impact of age may vary in different contexts [14].

Moreover, the establishment of trust in AVs encompasses both cognitive and emotional factors, with emotions serving as the primary determinant of trusting behavior [15]. However, less

Received 11 March 2024; revised 2 May 2024 and 3 July 2024; accepted 28 August 2024. Date of publication 13 September 2024; date of current version 20 November 2024. This work was supported by National Science Foundation under Grant 2138274. This article was recommended by Associate Editor Shi Cao. *(Lilit Avetisyan and Jackie Ayoub contributed equally to this work.) (Corresponding author: Feng Zhou.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Board at the University of Michigan Application No. HUM00215101.

Lilit Avetisyan, Jackie Ayoub, and Feng Zhou are with the Department of Industrial and Manufacturing Systems Engineering, The University of Michigan, Dearborn, MI 48128 USA (e-mail: lilita@umich.edu; jyayoub@umich.edu; fezhou@umich.edu).

X. Jessie Yang is with the Industrial and Operations Engineering, and Robotics, University of Michigan, Ann Arbor, MI 48109 USA (e-mail: xijyang@umich.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/THMS.2024.3452411.

Digital Object Identifier 10.1109/THMS.2024.3452411

attention has been paid to the nuanced and dynamic emotional experiences users have while interacting with AVs with few examples we found in the literature. For example, Avestian et al. [11] found that trust was significantly correlated with emotions that a high level of trust significantly improved participants' positive emotions, and vice versa. However, emotions might shape trust levels in complex ways beyond binary trust/distrust judgments, especially when different personality traits. Thus, additional research is required to gain insight into the intricate nuances of how various emotions impact trust in AVs and how these emotional responses may interact with other factors, such as the reliability of the system, driver personality traits, and prior experiences with AVs.

Another research gap in previous studies (e.g., [16]) pertains to the investigation of individual differences, such as dispositional trust and personality, in relation to trust dynamics. However, these studies primarily focused on capturing trust levels before and after experiments or through cross-sectional surveys, providing only a snapshot of trust dynamics. To gain a comprehensive understanding, it is essential to explore how individual factors influence trust dynamics across various trust profiles. Previous research demonstrated the existence of different trust dynamics, including oscillators, disbelievers, and Bayesian decision makers, in the context of human–robot interaction [17], [18] and human–automation interaction [19]. Therefore, it is imperative to examine the associations between different individual factors and trust dynamics in the domain of human–AV interaction.

In this study, we explored contextualized trust profiles in relation to varying AV reliability, focusing on factors such as personality, emotions, dispositional tendencies, initial perceptions, and real-time trust adjustments. A contextualized trust profile emphasizes the dynamic nature of trust of individual characteristics and a profile that depends on the specific driving situation and the AVs' behavior in those contexts. We explored the trust formation process, which laid the foundation for investigating long-term trust evolution, resulting from extensive human–AV interaction experiences, such as transitions between trust categories over time.

First, we employed a user segmentation approach in our user research to assist designers in understanding the behaviors of different contextualized trust profiles. This approach is particularly valuable for comprehending user trust in AVs as it enables designers to identify specific factors, particularly those related to individual differences, that influence trust, and subsequently, design interventions targeting these factors.

Second, building upon previous studies that demonstrated how dispositional trust and initial learned trust encompass numerous individual factors such as age, gender, culture, driving experience, and knowledge in automation [14], we minimized the number of factors by investigating dispositional and initial learned trust.

Third, we utilized a data-driven methodology using machine learning models, including K-means clustering to identify trust profiles, multinomial logistic regression with shapley additive explanations (SHAP) to validate these profiles [20], [21], and statistical comparisons among the identified profiles. These insights enable designers to develop trust-oriented solutions that effectively address relevant factors.

Overall, this research contributes to the understanding of trust dynamics in AVs and provides practical guidance for designers to enhance trust through tailored design interventions.

II. Related Work

In order to understand trust in AVs, it is important to understand the factors that affect trust. Researchers have identified a wide range of factors that impact trust in AVs. Hoff and Bashir [14] proposed a taxonomy of these factors based on three types of trust, namely dispositional, learned, and situational trust. In the context of driver–AV interaction, dispositional trust represented drivers' tendency to trust AVs, including typical factors such as age, gender, culture, and personality; learned trust represents the drivers' assessment of the AV based on their past experience (e.g., preexisting knowledge about the AV) or current interaction (e.g., reliability and performance of the AV) with the AV; situational trust is dependent on the interaction between drivers and the automation in specific contexts, including external environment of the interaction between the driver and the AV (e.g., the complexity and risks associated with the task) and internal characteristics of the driver (e.g., the emotional states and cognitive workload of the driver).

It is crucial to investigate individual factors that influence dispositional trust. For instance, Robert et al. [22] found that participants in high-context cultures, such as East Asia (including China, South Korea, and Japan) exhibited greater trust in AVs compared to individuals from low-context cultures (e.g., Western Europe and US) when explanations were provided. Personality was also identified to be a vital factor on trust. Chien et al. [23] demonstrated that higher levels of agreeableness and conscientiousness in personality traits were associated with increased initial trust in automation, with agreeableness and conscientiousness being two of the dimensions in the Big Five model of personality. However, experimental examination of these factors can often be challenging due to contextualization and potential interactions with other factors, such as age and driving experience. This can result in inconsistent findings, as observed in the effects of age on trust in AVs [12], [13]. Thus, it is crucial to investigate multiple individual factors concurrently to gain a comprehensive understanding of their overall impact on trust in AVs.

For learned trust, numerous studies examined preexisting knowledge and experience of AVs through surveys and current interaction through experimental studies. For example, Ayoub et al. [10] found that knowledge of AVs generally increased trust in AVs while experience in driving decreased trust in AVs. As mentioned previously, during interaction with AV, a high level of system transparency, reliability, and performance generally increased trust in AVs [4], [5], [6]. However, it is crucial to consider the temporal aspect of trust development, particularly during extended interaction periods with AVs. To explore this, Bhat et al. [17] conducted a study involving sequential decision-making tasks and employed a clustering model to identify distinct trust dynamics and different types of trust profiles.

For situational trust, both external variability and internal factors should be considered to understand trust in AVs. For example, Azevedo-Sa et al. [6] found that trust in AVs increased over time when internal risk was low (i.e., system reliability was high) while external risks (visibility in driving) did not impact trust significantly.

While the effects of many other factors on trust were investigated, less is known about the effects of emotions on trust in AVs. Previous research has suggested that positive emotions may foster trust due to their association with feelings of safety and security [24]. For instance, Du et al. [25] discovered that positive emotions improved takeover performance in AVs, subsequently leading to increased trust. Ayoub et al. [10] demonstrated a positive correlation between the feeling of excitement and participants' trust in AVs. Conversely, negative emotions have been found to diminish trust, even when unrelated to trust-related decisions [26]. Myers et al. [24] investigated the influence of three negative emotions (anger, guilt, and anxiety) on trust and found that negative emotions with low certainty appraisals (e.g., anxiety) reduced trust, whereas those with high certainty (e.g., anger and guilt) had no discernible effects on trust. In addition, Ayoub et al. [1] revealed that negative emotions such as concerns and worries decreased parents' trust in automated school buses. Thus, it is imperative to examine the effects of emotions and their role in identifying contextualized trust profiles in the context of AVs.

Overall, previous studies have highlighted several key factors that influence the formation of initial trust profiles in AVs. Dispositional factors like age, gender, culture, personality traits, and general propensity to trust technology seem to predispose individuals toward higher or lower baseline trust levels. These initial trust tendencies are then calibrated through learned experience with an AV. However, emotional responses and perceived risk in specific situational contexts further modulate trust levels dynamically, going beyond objective performance assessments. A key gap remains in understanding how the complex interplay between an individual's dispositional tendencies, learned system judgments, and emotional reactions in specific contexts formulates distinct trust profiles among people. Some individuals may exhibit consistent, goal-oriented trust focused on the long-term benefits of AVs. Others react more volatilely, with trust oscillating sensitively to the AV's short-term successes and failures. While some maintain persistent skepticism, requiring a fundamental understanding of the AV's underlying processes before trusting. Therefore, examining how individual differences across these factors interact to form distinct trust profiles is crucial for designing personalized experiences that can effectively calibrate and enable appropriate trust for widespread AV adoption.

III. Methodology

Fig. 1 provides an overview of the methodology in this study and the details are described in the following.

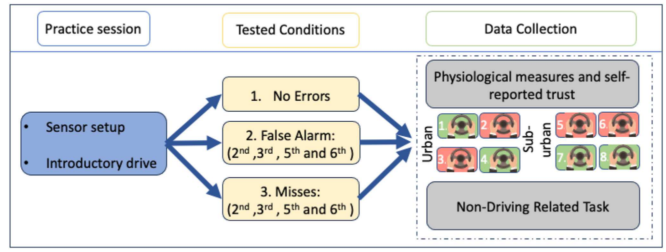

Fig. 1. Framework of the proposed study. In data collection, the green numbers indicate true TORs, while the red numbers represent false alarms or missed TORs.

Fig. 2. Experiment setup.

A. Participants

In this study, we recruited 74 university students, each of whom received a compensation of $25 for their participation, involving a duration of approximately one hour. Four participants were excluded from further analysis due to missing data. Data from the remaining 70 participants (mean age = 21.3, SD = 3.0; age range: 18 to 33; 32 female and 38 male participants) were utilized for subsequent analysis. All participants satisfied the requisite criterion of holding a valid driver's license with normal or corrected-to-normal vision.

B. Apparatus and Stimuli

The study employed a desktop-based driving simulator from Realtime Technologies Inc. (RTI, MI, USA) for data collection, as illustrated in Fig. 2. The simulation setup comprised three LCD monitors, integrated with a Logitech driving kit, and two touchscreens (i.e., a tablet and phone), located on the participant's right side, for the NDRT and trust rating entry. The NDRT involved a custom-designed Tetris game, implemented using the PyGame library in Python. Participants were required to drag tiles to navigate the game, which they could pause to respond to takeover requests (TORs), and then, resume from the same position afterward. Trust was evaluated using a questionnaire created with Qualtrics (Provo, UT, USA, www.qualtrics.com) on a mobile phone, following previous studies [27], [28].

The driving simulation was designed to simulate SAE Level 3 automation. To activate the automated mode, participants were required to depress a red button situated on the steering wheel. Upon initiation of the automated mode, an auditory message, "Automated mode engaged," would be emitted, after which the car would proceed along a predefined route while maintaining a speed of 35 mi/h. During the automated driving, participants

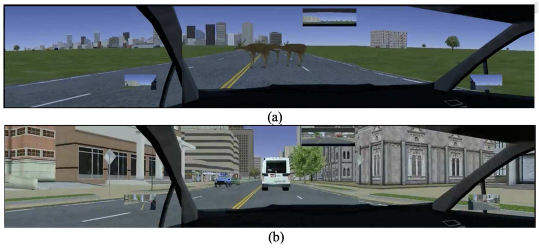

Fig. 3. Examples of takeover event in (a) suburban area with dears ahead and (b) urban area with a bus sudden stop ahead.

were directed to perform a NDRT with an auditory request "Please start the secondary task," which involved playing the Tetris game on a tablet. In the event of a TOR, participants would be alerted by an auditory warning of "Takeover" and the automated mode would be promptly disengaged to facilitate the participant's control of the vehicle. In instances where participants failed to resume control of the vehicle within the designated time frame (i.e., 7 s). including scenarios involving control and false alarm (FA) situations where TOR was initiated, as well as the missed condition where no TOR was made and participants failed to notice the road hazard themselves, an auditory warning ("Emergency Stop") would sound, triggering an immediate emergency stop to prevent potential collisions .

C. Experimental Design

In this study, we employed a between-subjects design in which participants were assigned to one of the three conditions: control, false alarm, or miss. In the control condition, all eight TORs were valid. In the false alarm condition, four TORs (the first, fourth, seventh, and eighth) were valid, and the other four TORs (the second, third, fifth, and seventh) were triggered unnecessarily, as the road was clear. Similarly, in the miss condition, four TORs (the first, fourth, seventh, and eighth) were valid; whereas in the other four TORs (the second, third, fifth, and sixth), there was an obstacle on the road, but AV did not detect it or initiate a TOR. This approach was adopted to evaluate the impact of different levels of automation performance on trust, as previous studies indicated that both false alarms and misses reduced trust in automated systems [29]. During the simulation, standard roadway features were used to signify the occurrence of a TOR event, such as the presence of deer, bicyclists, pedestrians, construction zones, vehicle sudden stops, buses with sudden stops, and police vehicles on the shoulder (see Fig. 3). The TOR events were equally distributed between rural and urban areas, with four TOR events taking place in each location. The order of the rural and urban scenarios was alternated to minimize potential order effects. To eliminate potential learning effects leading to predictable expectations, we varied the distances at which TORs occurred, with time intervals ranging from 2 to 4 min.

D. Experimental Procedure

Upon arrival, participants were asked to complete a consent form and an online demographic survey. After that, participants received an introduction and watched a short video about the tasks they were required to do. Then, participants completed an online survey that consisted of personality information, propensity to trust AVs, and initial learned trust. The Big Five Model [30] was used to examine different personality traits of the participants. We evaluated participants' personalities using the Big Five Inventory scale (BFI-10) [31] in five-point Likert scales, measuring character traits across five dimensions: extraversion (i.e., sociable and reversing reserved), agreeableness (i.e., trusting and reversing carper), conscientiousness (i.e., meticulous and reversing lazy), neuroticism (i.e., nervous and reversing relaxed), and openness to experience (i.e., imaginative and reversing artistic). The dispositional trust was measured using a six-item trust scale proposed by Merritt et al. [32]. The statements (see Table III) were rated with seven-point Likert scales. The initial learned trust was assessed with ten items (see Table III) using seven-point Likert scales. Subsequently, the participants underwent a training session to familiarize themselves with the driving simulator and experimental protocol. They were required to assess the automated mode functionality of the vehicle and instructed to maintain vigilance and resume control when necessary. Furthermore, they were informed about potential system failure scenarios, including instances of failing to detect obstacles (in the miss condition) and false alarms of TORs (in the false alarm condition). There were two drives, including urban and suburban with approximately 15 min each. The entire experiment was completed within a duration of approximately 60 min. To monitor their trust levels dynamically, participants were prompted to respond to a single-item trust rating, ranging from 0 to 10, every 25 s [27], [28], with the aim of avoiding any excessive mental or emotional stress [33]. In addition, they were asked to rate their trust levels after completing the driving simulation [27]. Finally, participants rated their anticipated emotional responses to AVs using a seven-point Likert scale that consisted of 19 emotion items, including disdainful, scornful, contemptuous, hostile, resentful, ashamed, humiliated, confident, secure, grateful, happy, respectful, nervous, anxious, confused, afraid, freaked out, lonely, and isolated, which were used to investigate trust in human–machine automation previously [11], [34].

IV. Results

A. Identifying Contextualized Trust Profiles

We attempted to identify contextualized trust profiles using a clustering method based on various measures recorded during the experiment, taking into account factors such as personality, emotions, dispositional, initial learned, and real-time dynamic trust, as well as the performance of the AV (i.e., control, false alarm, and misses). To do this, 48 features from these measures were first normalized between 0 and 1, and then, used as input for the K-means clustering model within the Azure Machine Learning environment. Through iterative optimization processes in the environment, the optimal number of clusters was determined to be 3. To validate this number, we employed the uniform manifold approximation and projection for dimension reduction (UMAP) method [35], which compared to the conventional approaches could reveal hidden structures in our complex dataset. The UMAP-transformed data in the new dimension confirmed the presence of three distinct clusters. As a result, three clusters were formed and named as follows:

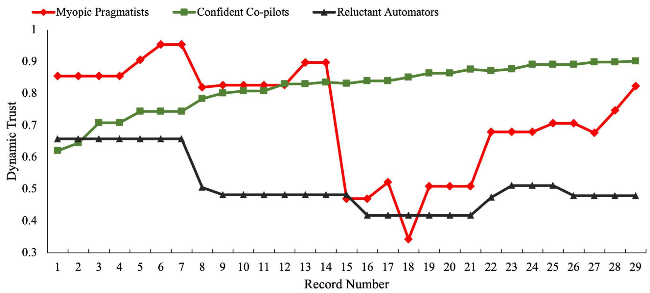

Fig. 4. Dynamic trust evaluation of three types of contextualized trust profiles during the experiment where the *x*-axis is the ordinal number of trust ratings in the experiment. Note the number of points of each participant were different, and we aggregated them by the order of their self-reported trust to show the dynamics of trust of three models.

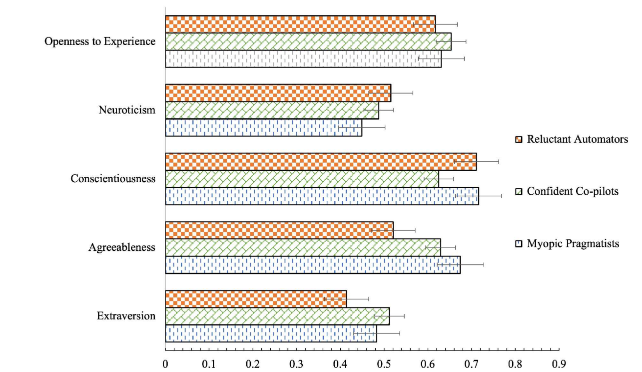

Fig. 5. Mean and standard deviation (SD) of Big Five factors (dimensions) of personalities among contextualized trust profiles.

1) *Myopic pragmatists*, consisting of 23 participants (8 female and 15 male), who were driven by the real-time performance of automation and could change their behaviors based on recent fluctuations. They were sensitive to both AV successes and failures, demonstrating a more watchful and cautious approach. This group was slower to fully relinquish control to the AV, often requiring consistent positive reinforcement to maintain trust;
2) *Confident copilots*, consisting of 31 participants (13 female and 18 male), who demonstrated consistently high trust in AVs, viewed the technology favorably, and maintained positive emotions even after experiencing system errors. These drivers prioritized the overall convenience and benefits of AVs, displaying resilience to occasional setbacks.
3) *Reluctant automators*, consisting of 16 participants (ten female and six male), characterized by low baseline trust and a preference for maintaining control. These drivers were hesitant to cede agency to AVs, focusing on understanding the vehicle's decision-making processes before extending trust. They often exhibited more pronounced negative emotional responses (e.g., fear and anxiety), preferring to rely on their own driving skills due to a desire for transparency and control.

The following subsections detail the findings of each group across trust, personality, and emotions dimensions.

B. Trust

1) Dynamic, Dispositional, and Initial Learned Trust: Examining the contextualized trust profiles, we found that first profile (i.e., *myopic pragmatists*) corresponded to drivers who exhibited a moderate high level of dynamic trust (Mean $=$ 0.61 and SD $= 0.23$), but reported high levels of dispositional (Mean $= 0.67$ and SD $= 0.19$) and initial learned trust (Mean $=$ 0.65 and SD $= 0.22$). Despite the fact that *myopic pragmatists* reported the highest levels of dispositional and initial learned trust, their trust level were dynamically evolving based on their most recent experiences resulting in rapid oscillation (see Fig. 4).

The *confident copilots* contextualized trust profile consisted of drivers who showed a higher level of dynamic trust (Mean $= 0.69$ and SD $= 0.15$) during the experiment, but reported moderate levels of dispositional (Mean $= 0.53$ and SD $= 0.22$) and initial learned trust (Mean $= 0.55$ and SD $= 0.21$). In this model, the dynamic trust showed an increasing trend as participants rated their trust considering their previous knowledge and experience besides the most recent ones happening during the experiment

The last profile (i.e., *reluctant automators*) included drivers who exhibited the lowest level of dynamic trust (Mean $= 0.45$ and SD $= 0.26$), as well as low levels of dispositional (Mean $=$ 0.40 and SD $= 0.24$) and initial learned trust (Mean $= 0.42$ and SD $= 0.24$). The individual data observations of *reluctant automators* showed that their levels of dynamic trust either decreased or remained low throughout the experiment. Also, drivers in this model tended to be carping rather than being optimistic about the situation and the AV's overall performance.

In examining dynamic trust changes statistically, the results of one-way analysis of variance (ANOVA) showed significant differences among contextualized trust profiles. Specifically, *reluctant automators* exhibited significantly distinct trust levels in urban and suburban areas. In the urban context, *reluctant automators* had lower trust levels compared to other contextualized trust profiles ($p = 0.000$), while in suburban areas, this difference was significant compared to *confident copilots* ($p = 0.017$).

Regarding dispositional and learned trust, the results of one-way ANOVA showed that there was a significant difference between three contextualized trust profiles for all the factors (see Table III) measuring dispositional and initial learned trusts ($p < 0.05$), except one from initial learned trust that measured trust in delegating control task to AV when driving was boring ($p = 0.08$).

C. Personality and Emotions

1) Personality: We conducted a one-way ANOVA test to compare these dimensions across the contextualized trust profiles and found a significant difference in agreeableness ($F(2, 67) = 2.452, p = 0.033$). The post hoc analysis with Tukey–Kramer test showed that *myopic pragmatists* had significantly higher levels of agreeableness compared to *reluctant automators* (see Fig. 5). In particular, a significant difference was

TABLE I
MEAN, STANDARD DEVIATION (SD), AND p-VALUE OF INDIVIDUAL ITEMS IN PERSONALITY

Feature	*Myopic pragmatists*	*Confident copilots*	*Reluctant automators*	*p-value*
Reserved	0.51 (0.27)	0.49 (0.36)	0.65 (0.31)	0.085
Trusting	0.77 (0.21)	0.65 (0.32)	0.67 (0.34)	0.305
Lazy	0.46 (0.28)	0.53 (0.30)	0.47 (0.34)	0.622
Relaxed	0.66 (0.21)	0.60 (0.29)	0.53 (0.34)	0.344
Artistic	0.51 (0.30)	0.52 (0.33)	0.48 (0.37)	0.950
Sociable	0.74 (0.22)	0.75 (0.25)	0.58 (0.36)	0.101
Carper ***	0.39 (0.28)	0.38 (0.28)	0.62 (0.25)	0.000
Meticulous	0.87 (0.15)	0.78 (0.25)	0.89 (0.13)	0.127
Nervous	0.57 (0.26)	0.57 (0.27)	0.56 (0.28)	0.991
Imaginative	0.76 (0.19)	0.82 (0.22)	0.72 (0.27)	0.291

Carper of reluctant automators was significantly higher than that of myopic pragmatists and confident copilots.

TABLE II
MEAN, STANDARD DEVIATION (SD), AND p-VALUE OF EMOTIONS

Feature	***Myopic Pragmatists***	***Confident Co-pilots***	***Reluctant Automators***	**p-value**
Lonely ***	0.18 (0.22) $_{a}$	0.06 (0.17) $_{ab}$	0.44 (0.35) $_{c}$	0.000
Isolated ***	0.20 (0.22) $_{a}$	0.08 (0.17) $_{ab}$	0.44 (0.35) $_{c}$	0.000
Resentful ***	0.30 (0.27) $_{a}$	0.10 (0.20) $_{b}$	0.34 (0.31) $_{ac}$	0.003
Humiliated ***	0.30 (0.30) $_{a}$	0.06 (0.12) $_{b}$	0.36 (0.40) $_{ac}$	0.000
Ashamed ***	0.33 (0.32) $_{a}$	0.06 (0.13) $_{b}$	0.39 (0.39) $_{ac}$	0.000
Hostile ***	0.37 (0.27) $_{a}$	0.10 (0.25) $_{b}$	0.56 (0.38) $_{ac}$	0.000
Disdainful ***	0.41 (0.22) $_{a}$	0.09 (0.18) $_{b}$	0.39 (0.28) $_{ac}$	0.000
Afraid ***	0.43 (0.32) $_{a}$	0.20 (0.24) $_{b}$	0.50 (0.23) $_{ac}$	0.000
Scornful ***	0.45 (0.35) $_{a}$	0.13 (0.29) $_{b}$	0.38 (0.32) $_{ac}$	0.001
Freaked out ***	0.45 (0.32) $_{a}$	0.06 (0.14) $_{b}$	0.40 (0.22) $_{ac}$	0.000
Grateful	0.54 (0.23)	0.62 (0.21)	0.48 (0.20)	0.080
Nervous ***	0.54 (0.25) $_{a}$	0.33 (0.26) $_{b}$	0.64 (0.20) $_{ac}$	0.000
Confused ***	0.56 (0.24) $_{a}$	0.24 (0.26) $_{b}$	0.68 (0.18) $_{ac}$	0.000
Contemptuous ***	0.57 (0.35) $_{a}$	0.23 (0.33) $_{b}$	0.52 (0.27) $_{ac}$	0.001
Secure ***	0.57 (0.23) $_{a}$	0.75 (0.14) $_{b}$	0.46 (0.25) $_{ac}$	0.000
Confident ***	0.59 (0.25) $_{a}$	0.72 (0.12) $_{b}$	0.55 (0.15) $_{ac}$	0.003
Happy ***	0.59 (0.27) $_{a}$	0.63 (0.20) $_{ab}$	0.34 (0.20) $_{c}$	0.000
Respectful ***	0.61 (0.27) $_{a}$	0.72 (0.21) $_{ab}$	0.45 (0.19) $_{ac}$	0.001
Anxious ***	0.66 (0.27) $_{a}$	0.38 (0.27) $_{b}$	0.70 (0.24) $_{ac}$	0.000

Note: The models that share the same letter (i.e., a, b, or c) are not statistically different.

found in the item "I see myself as someone who tends to find fault with others" statement ($p = 0.015$) where *reluctant automators* had higher ratings than *confident copilots* and *myopic pragmatists* (see Table I). However, there were no significant differences in the other four personality characteristics (i.e., extraversion, conscientiousness, neuroticism, and openness to experience). This result was consistent with the SHAP model outcome, as only the "carper" feature, which represented the aforementioned questions related to personality, was selected as an important feature in the model.

2) Emotions: The results of the one-way ANOVA test indicated that there were significant differences in all the emotions among the three contextualized trust profiles, except "Grateful," as detailed in Table II. To understand the underlying structure of emotions associated with each contextualized trust profile, we conducted an exploratory factor analysis and identified three subsets that combined the correlated emotions, i.e., *resentfully aversion:* disdainful, scornful, contemptuous, hostile, resentful, ashamed, and humiliated; *happily acceptance:* confident, grateful, secure, happy, respectful, not lonely, and not isolated; and *nervously fear:* nervous, anxious, confused, afraid, and freaked out. The results showed that *confident copilots* had significantly lower scores for emotions related to resentfully aversion compared to the other contextualized trust profiles ($p = 0.000$). They also had significantly higher scores for confidence and security in the happily acceptance category compared to *myopic pragmatists*, and higher scores for happiness, gratitude, and respect compared to *reluctant automators*. For nervously fear, *confident copilots* showed significantly lower scores than the other contextualized trust profiles ($p = 0.000$). Moreover,

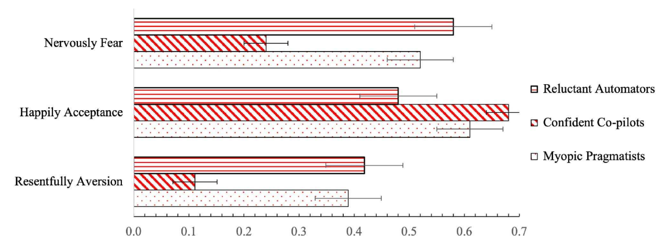

Fig. 6. Mean and standard deviation (SD) of three emotion categories among contextualized trust profiles.

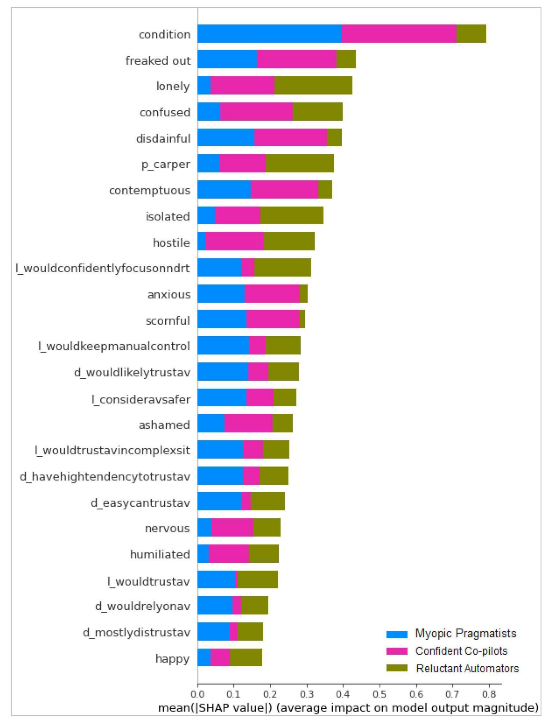

Fig. 7. Summary of SHAP values where the features are ordered from the highest to the lowest effect on the prediction. In the figure, the prefix "p_" in features indicates personality-related items, "d_" indicates dispositional trust items, and "l_" indicates learned trust items.

myopic pragmatists had significantly higher scores for happiness ($p = 0.016$) compared to *reluctant automators* (see Fig. 6).

D. Validating Contextualized Trust Profiles

In order to validate the contextualized trust profiles, we used the cluster membership as the ground truth and tested the accuracy of the clustering model by training supervised machine learning models with all the 48 features. We used an Azure automated ML experiment job to identify the best machine learning model in Azure automated ML experiment by fine-tuning the hyperparameters. The multinomial logistic regression model was found to perform the best across a large number of models, including XGBoost, LightGBM, Random Forest, etc. The results showed that the model was able to predict the contextualized trust profile with F1-score = 0.90.

We also used SHAP explainer to interpret the model and understand the importance of individual features in the model's

TABLE III
MEAN, STANDARD DEVIATION (SD), AND p-VALUE OF DISPOSITIONAL AND INITIAL LEARNED TRUST

	Feature	*Myopic Pragmatists*	*Confident Co-pilots*	*Reluctant Automators*	p-value
Dispositional	I usually trust self-driving vehicles until there is a reason not to	0.81 (0.16) $_a$	0.67 (0.20) $_b$	0.45 (0.26) $_c$	0.000
	For the most part, I distrust self-driving vehicles	0.24 (0.21) $_a$	0.39 (0.21) $_b$	0.57 (0.25) $_c$	0.000
	In general, I would rely on a self-driving vehicle to assist me	0.78 (0.18) $_a$	0.61 (0.19) $_b$	0.41 (0.27) $_c$	0.000
	My tendency to trust self-driving vehicles is high	0.78 (0.19) $_a$	0.53 (0.21) $_b$	0.35 (0.24) $_c$	0.000
	It is easy for me to trust self-driving vehicles to do their job	0.78 (0.16) $_a$	0.56 (0.21) $_b$	0.42 (0.25) $_{bc}$	0.000
	I am likely to trust self-driving vehicles even when I have little knowledge about it	0.66 (0.27) $_a$	0.42 (0.28) $_b$	0.23 (0.19) $_{bc}$	0.000
Initial learned	I would feel safe in a self-driving vehicle	0.80 (0.11) $_a$	0.65 (0.16) $_b$	0.46 (0.22) $_c$	0.000
	The self-driving vehicle system provides me with more safety	0.70 (0.25) $_a$	0.45 (0.20) $_b$	0.39 (0.22) $_{bc}$	0.000
	I would rather keep manual control of my vehicle than delegate it to the self-driving vehicle system on every occasion	0.39 (0.30) $_a$	0.57 (0.29) $_{ab}$	0.76 (0.18) $_{bc}$	0.000
	I would trust the self-driving vehicle system decisions	0.76 (0.14) $_a$	0.65 (0.13) $_{ab}$	0.43 (0.25) $_c$	0.000
	I would trust the self-driving vehicle system capacities to manage complex driving situations	0.64 (0.24) $_a$	0.42 (0.21) $_b$	0.33 (0.28) $_{bc}$	0.000
	If the weather conditions were bad, I would delegate the driving task to the self-driving vehicle system	0.45 (0.30) $_a$	0.35 (0.25) $_{ab}$	0.22 (0.19) $_{bc}$	0.030
	Rather than monitoring the driving environment, I could focus on other activities confidently	0.61 (0.24) $_a$	0.48 (0.29) $_{ab}$	0.26 (0.24) $_c$	0.001
	If driving was boring for me, I would rather delegate it to the self-driving vehicle system than do it myself	0.70 (0.27)	0.67 (0.21)	0.52 (0.28)	0.081
	I would delegate the driving to the self-driving vehicle system if I was tired	0.79 (0.24) $_a$	0.74 (0.20) $_{ab}$	0.59 (0.29) $_{bc}$	0.042
	I would trust the self-driving vehicle	0.76 (0.15) $_a$	0.55 (0.20) $_b$	0.30 (0.26) $_c$	0.000

Note: The models that share the same letter (i.e., a, b, or c) are not statistically different.

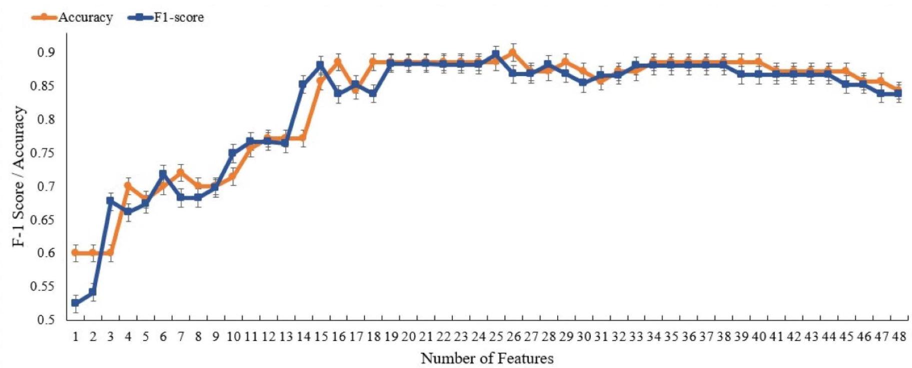

Fig. 8. Impact of individual features on overall performance of the model. During each iteration, a single feature was added to the model based on SHAP importance ranking. The error bars show the standard errors computed from fivefold cross validation.

decision-making process. Fig. 7 illustrates the feature importance rankings, from the most influential to the least influential on the prediction. The features of condition (miss, false alarm, or control), emotions of freaked out and confused were found to be the most significant ones in predicting the contextualized trust profiles, regardless of whether they had a positive or negative effect. The color coding indicates the feature importance in distinguishing between different contextualized trust profiles. We used a feature selection method to improve the model's performance while preserving the patterns and relationships in the data, as suggested by Ayoub et al. [20], [21]. We added one feature at a time based on the feature importance ranking from the SHAP explainer and validated the results with stratified fivefold cross validation. As shown in Fig. 8, the model reached the highest performance (F1-score $= 0.90$) using the top 25 features, i.e., condition, *emotions:* freaked out, lonely, confused, disdainful, contemptuous, isolated, hostile, ashamed, anxious, scornful, nervous, humiliated, and happy; *personality:* carper, *initial learned trust:* I could confidently focus on NDRT, I would keep manual control, I would consider AV more safe than manuals, I would trust in complex situations, and I would trust AV; and *dispositional trust:* I likely to trust AV without knowledge, I have high tendency to trust AV, I easily can trust AV, I would rely on an AV, and I mostly distrust AV. Overall, across all the 48 features, the F1-score ranged from 0.52 to 0.90, and the accuracy ranged from 0.60 to 0.90.

Even though the experiment condition was found to be the most influential feature in predicting these three contextualized trust profiles, it was not sufficient to determine the identified contextualized trust profiles alone. As shown in Fig. 9, there was no one-to-one mapping relationships between the participants in three conditions, including control, FA, and misses and those formed the three identified contextualized trust profiles. Also shown in Fig. 8, the condition alone only had F1-score $= 0.52$ and accuracy $= 0.60$ in predicting contextualized trust profiles. Other factors, such as emotions, personality, initial learned trust, and dispositional trust, also played important roles in determining these profiles.

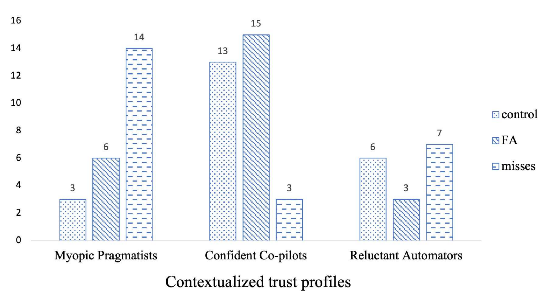

Fig. 9. Distribution of the clusters across the three conditions.

V. DISCUSSION

A. Implications

Our study identified three distinct contextualized trust profiles: *myopic pragmatists*, *confident copilots*, and *reluctant* *automators*. These profiles, predicted with good accuracy by the proposed model, demonstrate the complexity of how individuals react to automated driving experiences. Trust emerges as highly dynamic, shaped by the intersection of personality, initial biases, emotional responses, and ongoing AV performance assessment.

Myopic pragmatists are performance-driven and emotionally volatile. They exhibit trust fluctuations closely tied to real-time AV performance. Successes bolster trust, while errors trigger emotional responses such as fear and disappointment [11]. This sensitivity might be advantageous in certain situations, prompting quicker reactions when necessary. Consistent with our previous research, trust here focuses on the system's performance capabilities within specific contexts [4]. However, unlike previous findings, myopic pragmatists seemed to easily recover from low levels of trust caused by errors made by the AV [36]. This is also represented by their corresponding emotional responses associated with overall experience. Moreover, as agreeable individuals, myopic pragmatists tend to be more adaptable and cooperative, which may explain their willingness to adjust trust levels based on recent experiences.

In order to help mitigate such volatility, the design should aim to deliver a consistently positive experience. This can be achieved by providing reliable and consistent performance across a variety of situations, such as inclement weather, heavy traffic, and challenging road conditions. In this aspect, adaptive automation might help myopic pragmatists to calibrate their trust by adapting the level of automation based on the drivers' current trust level in AV and the system's past performance [37]. For instance, if the driver is confident and happy in the AV, the system can perform in conditional autonomy and require less intervention from the driver. On the other hand, if the driver is less confident in the AV resulting from previous negative experiences, the system can provide feedback about system status in order to increase the feelings of confidence, control, and safety. Furthermore, since negative experiences might stem from either excessive trust or unfamiliarity with the AV's capabilities, the AV can require driver intervention to familiarize them with its behavior, demonstrating its range of capabilities and response patterns to enhance the driver's understanding and trust in the system. Therefore, by integrating adaptive automation into the AV system, designers would help myopic pragmatists to calibrate their trust in the automation system, improving overall system performance and safety [37], [38].

Confident copilots show consistently a high level of trust, fueled by positive experiences and potentially downplaying occasional errors. They focus on the overall purpose of the AV and demonstrate greater emotional resilience even in the face of setbacks. While aligning with prior work highlighting purpose-driven trust [15], [39], their trajectory diverges from Bayesian decision makers [17], as they maintain a trend of relatively higher level of trust than stabilizing trust levels. From the emotion point of view, confident copilots had a more positive attitude toward AV performance and were less prone to fear or disappointment when encountering failures. Specifically, they felt a greater sense of safety and were less vulnerable to false alarms, even if the TORs were unfounded, in contrast to reluctant automators who regarded such incidents as severe system malfunctions.

In order to design for confident copilots, the design should aim to provide features that increase their confidence and comfort levels, such as clear and intuitive interfaces, well-defined communication protocols, and robust safety features that effectively communicate the system's status and capabilities. They also tend to relate to the purpose of trust [15], [39] and believe that automation could help improve efficiency, reduce errors, and enhance safety by delegating their tasks to machines and freeing up their own cognitive and physical resources to focus on other tasks. This could be not limited to automated driving but in various domains, such as healthcare, aviation, and manufacturing.

However, confident copilots exhibited significantly more positive emotions, which might impact their tendency to take risks [40] and potentially lead to overtrust in AVs. To address this issue, the design should incorporate clear and transparent communication that explains the system's capabilities and limitations by clearly showing the limitations of the system [41], and include educational materials to help them understand how the technology works. For example, the system should be designed to clearly communicate when the system would request the driver to take over control.

Reluctant automators possess low baseline trust, likely rooted in a need to understand the AV's underlying processes for trust to emerge. This aligns with process-based trust [39] and the group of participants identified in [17] who generally had a low level of trust in automation. Despite the AV's capabilities, they maintain negative emotional responses (e.g., fear and nervousness) and insist on having control over driving. Their persistent negative emotions and nonagreeable personality traits create a barrier to trust that goes beyond simply enhancing AV performance or providing more information. To build trust with this group, strategies may need to focus more on addressing underlying emotional and personality factors, potentially through long-term exposure and gradual introduction of autonomy.

Regarding predictive models, while Bhat et al.'s model [17] successfully predicted the Bayesian decision makers and disbelievers, our model, which incorporates emotions, personality traits, and both learned and dispositional trust, predicts all three trust profiles more accurately. This highlights the need to consider a broader range of variables beyond dynamic trust in predicting contextualized trust profiles.

Given the distinct emotional profiles associated with each contextual trust profile, future AV designs could benefit from incorporating emotion recognition and response mechanisms. For example, detecting increased anxiety in reluctant automators could trigger more detailed explanations of the AV's decision-making process. For myopic pragmatists, the system could provide reassurance and positive feedback after successful maneuvers to help stabilize their emotional state and trust levels.

While our study identified significant differences in Agreeableness among trust groups, future research should investigate how other personality traits might interact with emotional responses to influence trust in AVs. In addition, these emotional and personality differences may be influenced by cultural factors. Cross-cultural studies could explore how these contextualized trust profiles manifest in different societies, potentially revealing new insights into the universality or cultural specificity of trust formation in AVs.

The significant role of AV performance in the trust formation process is integral to establishing and sustaining user trust over time. The most important aspect is to improve the vehicle reliability itself [4], such as rigorous testing and refinement of perception and control systems to ensure handling complex scenarios and takeover transitions. However, human factors issues also play an important role, such as transparent communication of the situation awareness [8], capabilities, and limitations so that users have a realistic understanding of what an AV perceives, can and cannot do in different scenarios. Previous studies showed that people exhibit a propensity to accept automation malfunctions if they are able to promptly regain manual control, thus mitigating the risk of potential accidents [42]. However, the nuances of transparency requirements hinge on the initial trust profile of the driver. Hence, adaptive transparency mechanisms are needed to further facilitate driver trust and improve their driving performance [43], [44].

Despite the success of the audio alert in prompting participants to take over control, ensuring a seamless transition requires attention to other interfaces. If the vehicle is designed so that they can easily understand what the system is doing and why it is doing it through clear and intuitive communication as feedback, this would be helpful [8]. For example, by explaining "why" and "what will" information with speech and augmented reality during the takeover process, participants reported it to be easy to use and accept SAE Level 3 vehicles [7]. We should also consider predictability of the system so that the system's behavior aligns with user expectations and how well it performs in different scenarios. For example, when participants received explanations about the vehicle's behavior ahead of the time and their possible projection in the future, they had better situation awareness of the driving scenarios and trust in automated driving [8]. These two strategies can help drivers to understand the system's behavior and performance, which in turn can enhance their learned trust and confidence in the system.

B. Limitations

This study has several limitations that require further investigation in future research. First, the study was conducted using a low-fidelity experimental setup utilizing a desktop driving simulator with its reliance on only an auditory alert of TORs. We recognize the need to refine and enhance the realism of our experimental setup to better align with the complexities of level 3 automation systems. The sample size of 70 participants was relatively small and we only considered a limited number of factors to identify contextualized trust profiles. Moreover, the study sample was composed primarily of university students, which resulted in a homogeneous sample, regarding age, education, driving experience, and knowledge about AVs. To overcome these limitations, future studies should be conducted in higher fidelity experimental settings and with a larger and more diverse sample size. More external factors should be included, such as time constraints, perceived risks, complexity of NDRTs, and situation awareness [8], [14], [45].

Second, our clustering approach aimed to discover natural groupings based on shared patterns across various trust dimensions. This enabled us to identify potential models without any preexisting assumptions. We then used multinomial logistic regression to confirm these clusters and see whether the extracted models were predictable based on the original features. In addition, we utilized SHAP to understand the relative importance of these features, providing insights for tailoring future AV experiences to these different user models. While we acknowledge that this validation may have limitations due to the use of the same dataset, both methods offer complementary strengths. Clustering reveals patterns within our sample, while regression hints at broader generalizability. Ideally, validation would involve a completely separate dataset. However, given our sample size constraints, this will be a crucial focus for future studies to assess the robustness of our identified models.

Finally, we only included a limited number of factors that formed the contextualized trust profiles. While these factors provided valuable insights into the dynamics of trust, they may not fully capture the complexity and diversity of trust. It is important to acknowledge that trust is a multifaceted construct influenced by various individual and contextual factors, which might not be fully accounted for in this study. Therefore, the generalizability of the study findings to other contexts and populations might be limited and future studies should include a wider range of trust indicators to enhance the external validity of the results.

VI. Conclusion

The purpose of our research was to examine the contextualized trust profiles in AVs and determine the underlying behavioral patterns of drivers that could be useful for designing profile-based systems. To accomplish this, we collected multidimensional data and clustered them into three contextualized trust profiles: *confident copilots*, *myopic pragmatists*, and *reluctant automators*. We used these profiles to build a logistic regression model that could predict the contextualized trust profile with accuracy of 0.89 and F1-score of 0.90. In addition, we used SHAP explainer to identify the most significant factors in the dataset that influenced the creation of contextualized trust profiles. Furthermore, we investigated the dynamic trust patterns among these profiles, as well as the associated initial learned trust, dispositional trust, emotional, and personality characteristics, based on which we discussed how to develop a system that can adjust AV's behavior based on the driver's contextualized trust profiles, and eventually promote their acceptance and adoption of such technologies.

References

[1] J. Ayoub and F. Zhou, "Investigating drivers' trust in autonomous vehicles' decisions of lane changing events," in *Proc. Hum. Factors Ergonom. Soc. Annu. Meeting*, 2020, vol. 64, no. 1, pp. 1274–1278.

[2] M. Jaussein, L. Lévêque, J. Deniel, T. Bellet, H. Tattegrain, and C. Marin-Lamellet, "How do non-driving-related tasks affect engagement under highly automated driving situations? A literature review," *Front. Future Transp.*, vol. 2, 2021, Art. no. 687602.

[3] L. Petersen, L. Robert, J. Yang, and D. Tilbury, "Situational awareness, driver's trust in automated driving systems and secondary task performance," *SAE Int. J. Connected Auton. Veh.*, vol. 2, no. 2, pp. 129–141, 2019, doi: 10.4271/12-02-02-0009.

[4] J. Ayoub, L. Avetisyan, M. Makki, and F. Zhou, "An investigation of drivers' dynamic situational trust in conditionally automated driving," *IEEE Trans. Human-Mach. Syst.*, vol. 52, no. 3, pp. 501–511, Jun. 2021.

[5] R. Luo, N. Du, and X. J. Yang, "Evaluating effects of enhanced autonomy transparency on trust, dependence, and human-autonomy team performance over time," *Int. J. Human–Comput. Interact.*, vol. 38, no. 18–20, pp. 1962–1971, 2022.

[6] H. Azevedo-Sa, H. Zhao, C. Esterwood, X. J. Yang, D. M. Tilbury, and L. P. Robert Jr, "How internal and external risks affect the relationships between trust and driver behavior in automated driving systems," *Transp. Res. C, Emerg. Technol.*, vol. 123, 2021, Art. no. 102973.

[7] N. Du, F. Zhou, D. Tilbury, L. P. Robert, and X. J. Yang, "Designing alert systems in takeover transitions: The effects of display information and modality," in *Proc. 13th Int. Conf. Automot. User Interfaces Interactive Veh. Appl.*, 2021, pp. 173–180.

[8] L. Avetisyan, J. Ayoub, and F. Zhou, "Investigating explanations in conditional and highly automated driving: The effects of situation awareness and modality," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 89, pp. 456–466, 2022.

[9] Y. Chen, Y. Li, W. Yan, and X. Wang, "The effect of gender and cultural differences on trust in automated vehicles," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2663, no. 2, pp. 140–148, 2017.

[10] J. Ayoub, X. J. Yang, and F. Zhou, "Modeling dispositional and initial learned trust in automated vehicles with predictability and explainability," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 77, pp. 102–116, 2021.

[11] L. Avetisian, J. Ayoub, and F. Zhou, "Anticipated emotions associated with trust in autonomous vehicles," in *Proc. Hum. Factors Ergonom. Soc. Annu. Meeting*, vol. 66, no. 1, pp. 199–203, 2022, doi: 10.1177/1071181322661002.

[12] R. Pak, A. C. McLaughlin, W. Leidheiser, and E. Rovira, "The effect of individual differences in working memory in older adults on performance with different degrees of automated technology," *Ergonomics*, vol. 60, no. 4, pp. 518–532, 2017.

[13] B. Schoettle and M. Sivak, "Motorists' preferences for different levels of vehicle automation: 2016," Univ. Michigan Sustain. Worldwide Transp., Ann Arbor, MI, USA, Tech. Rep. SWT-2016-8 2016.

[14] K. A. Hoff and M. Bashir, "Trust in automation: Integrating empirical evidence on factors that influence trust," *Hum. Factors*, vol. 57, no. 3, pp. 407–434, 2015.

[15] J. D. Lee and K. A. See, "Trust in automation: Designing for appropriate reliance," *Hum. Factors*, vol. 46, no. 1, pp. 50–80, 2004.

[16] S. M. Merritt and D. R. Ilgen, "Not all trust is created equal: Dispositional and history-based trust in human-automation interactions," *Hum. Factors*, vol. 50, no. 2, pp. 194–210, 2008.

[17] S. Bhat, J. B. Lyons, C. Shi, and X. J. Yang, "Clustering trust dynamics in a human-robot sequential decision-making task," *IEEE Robot. Automat. Lett.*, vol. 7, no. 4, pp. 8815–8822, Oct. 2022.

[18] Y. Guo and X. J. Yang, "Modeling and predicting trust dynamics in human–robot teaming: A Bayesian inference approach," *Int. J. Social Robot.*, vol. 13, no. 8, pp. 1899–1909, 2020.

[19] H. Chung and X. J. Yang, "Associations between trust dynamics and personal characteristics," in *Proc. IEEE 4th Int. Conf. Human- Mach. Syst.*, 2024, pp. 1–6.

[20] J. Ayoub et al., "Cause-and-effect analysis of ADAS: A comparison study between literature review and complaint data," in *Proc. 14th Int. Conf. Automot. User Interfaces Interactive Veh. Appl.*, 2022, pp. 139–149.

[21] J. Ayoub, N. Du, X. J. Yang, and F. Zhou, "Predicting driver takeover time in conditionally automated driving," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 9580–9589, Jul. 2022.

[22] L. Robert, "Personality in the human robot interaction literature: A review and brief critique," in *Proc. 24th Amer. Conf. Inf. Syst.*, Aug. 2018, pp. 16–18.

[23] S.-Y. Chien, K. Sycara, J.-S. Liu, and A. Kumru, "Relation between trust attitudes toward automation, Hofstede's cultural dimensions, and big five personality traits," in *Proc. Hum. Factors Ergonom. Soc. Annu. Meeting*, 2016, vol. 60, pp. 841–845.

[24] C. D. Myers and D. Tingley, "The influence of emotion on trust," *Political Anal.*, vol. 24, no. 4, pp. 492–500, 2016.

[25] N. Du et al., "Examining the effects of emotional valence and arousal on takeover performance in conditionally automated driving," *Transp. Res. C, Emerg. Technol.*, vol. 112, pp. 78–87, 2020.

[26] J. B. Engelmann, F. Meyer, C. C. Ruff, and E. Fehr, "The neural circuitry of affect-induced distortions of trust," *Sci. Adv.*, vol. 5, no. 3, 2019, Art. no. eaau3413.

[27] M. Desai, P. Kaniarasu, M. Medvedev, A. Steinfeld, and H. Yanco, "Impact of robot failures and feedback on real-time trust," in *Proc. IEEE 8th ACM/IEEE Int. Conf. Human-Robot Interact.*, 2013, pp. 251–258.

[28] J. Ayoub, L. Avetisian, X. J. Yang, and F. Zhou, "Real-time trust prediction in conditionally automated driving using physiological measures," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 12, pp. 14642–14650, Dec. 2023.

[29] V. L. Pop, A. Shrewsbury, and F. T. Durso, "Individual differences in the calibration of trust in automation," *Hum. Factors*, vol. 57, no. 4, pp. 545–556, 2015.

[30] M. Zuckerman, D. M. Kuhlman, J. Joireman, P. Teta, and M. Kraft, "A comparison of three structural models for personality: The big three, the big five, and the alternative five," *J. Pers. Social Psychol.*, vol. 65, no. 4, 1993, Art. no. 757.

[31] B. Rammstedt and O. P. John, "Measuring personality in one minute or less: A 10-item short version of the big five inventory in English and German," *J. Res. Pers.*, vol. 41, no. 1, pp. 203–212, 2007.

[32] S. M. Merritt, H. Heimbaugh, J. LaChapell, and D. Lee, "I trust it, but i don't know why: Effects of implicit attitudes toward automation on trust in an automated system," *Hum. Factors*, vol. 55, no. 3, pp. 520–534, 2013.

[33] S. Hergeth, L. Lorenz, R. Vilimek, and J. F. Krems, "Keep your scanners peeled: Gaze behavior as a measure of automation trust during highly automated driving," *Hum. Factors*, vol. 58, no. 3, pp. 509–519, 2016.

[34] T. Jensen, M. M. H. Khan, Y. Albayram, M. A. A. Fahim, R. Buck, and E. Coman, "Anticipated emotions in initial trust evaluations of a drone system based on performance and process information," *Int. J. Human–Comput. Interact.*, vol. 36, no. 4, pp. 316–325, 2020.

[35] L. McInnes, J. Healy, and J. Melville, "Umap: Uniform manifold approximation and projection for dimension reduction," 2018, *arXiv:1802.03426*.

[36] R. Parasuraman and D. H. Manzey, "Complacency and bias in human use of automation: An attentional integration," *Hum. Factors*, vol. 52, no. 3, pp. 381–410, 2010.

[37] N. Moray, T. Inagaki, and M. Itoh, "Adaptive automation, trust, and self-confidence in fault management of time-critical tasks," *J. Exp. Psychol., Appl.*, vol. 6, no. 1, 2000, Art. no. 44.

[38] D. B. Kaber, M. C. Wright, L. J. Prinzel III, and M. P. Clamann, "Adaptive automation of human-machine system information-processing functions," *Hum. Factors*, vol. 47, no. 4, pp. 730–741, 2005.

[39] R. C. Mayer, J. H. Davis, and F. D. Schoorman, "An integrative model of organizational trust," *Acad. Manage. Rev.*, vol. 20, no. 3, pp. 709–734, 1995.

[40] F. Zhou, Y. Ji, and R. J. Jiao, "Prospect-theoretic modeling of customer affective-cognitive decisions under uncertainty for user experience design," *IEEE Trans. Human-Mach. Syst.*, vol. 44, no. 4, pp. 468–483, Aug. 2014.

[41] Y. Jiang, X. Li, and Y. Yang, "The effects of information on overtrust in highly automated driving," *Appl. Ergonom.*, vol. 82, 2020, Art. no. 102936.

[42] M. Seet, J. Harvy, R. Bose, A. Dragomir, A. Bezerianos, and N. Thakor, "Differential impact of autonomous vehicle malfunctions on human trust," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 1, pp. 548–557, Jan. 2022.

[43] F. Hartwich, C. Hollander, D. Johannmeyer, and J. F. Krems, "Improving passenger experience and trust in automated vehicles through user-adaptive HMIs: "The more the better" does not apply to everyone," *Front. Hum. Dyn.*, vol. 3, 2021, Art. no. 669030.

[44] P. Wintersberger, F. Janotta, J. Peintner, A. Löcken, and A. Riener, "Evaluating feedback requirements for trust calibration in automated vehicles," *It- Inf. Technol.*, vol. 63, no. 2, pp. 111–122, 2021.

[45] J. Manchon, M. Bueno, and J. Navarro, "From manual to automated driving: How does trust evolve?," *Theor. Issues Ergonom. Sci.*, vol. 22, no. 5, pp. 528–554, 2021.