Human-Machine Interface Dynamics and Driver Trust Calibration in Semi-Autonomous Vehicles Utilizing Real-Time OverFeat Perception-Based Lane Keeping and Collision Warnings

Authors

Abey City

Date: August 24, 2025

Abstract

Semi-autonomous vehicle adoption faces a critical challenge: the discrepancy between system capabilities and driver trust calibration, manifesting as misuse (over-trust leading to complacency) or disuse (distrust leading to system rejection). Existing human-machine interface (HMI) designs inadequately address the dynamic nature of trust formation during real-world driving, particularly regarding perception system transparency. This study develops and validates a novel HMI framework integrating real-time OverFeat-based perception outputs (lane boundaries and obstacle detection) with trust-calibration mechanisms in semi-autonomous vehicles. Using a driving simulator experiment with 60 participants across varying automation levels, we measured trust calibration through behavioral (takeover response time, NDRT engagement) and self-report metrics. The OverFeat CNN architecture achieved 89.4% lane detection accuracy and 87.2% vehicle detection precision at >10 Hz processing speeds. Results demonstrate that real-time perception visualization significantly improves trust calibration ($p < .01$) compared to conventional black-box HMI designs, reducing takeover response time variability by 34%. The framework provides actionable guidelines for HMI designers and contributes to the theoretical understanding of trust dynamics in human-automation interaction.

Keywords: Trust Calibration, Human-Machine Interface, OverFeat CNN, Semi-Autonomous Vehicles, Lane Keeping, Collision Warning, Driver Behavior

1. Introduction

1.1 Background

Road traffic accidents claim approximately 1.19 million lives annually worldwide, with over 90% of fatalities attributed to driver error . Semi-autonomous driving systems (SAE Level 2-3) aim to mitigate these errors by delegating routine control to automation while maintaining driver supervisory responsibility. The effectiveness of these systems hinges on two interdependent factors: robust perception capable of reliably detecting lane boundaries and obstacles, and appropriate driver trust calibration that aligns reliance with actual system capabilities. Recent advances in deep learning have significantly improved perception performance in autonomous driving contexts. Convolutional Neural Networks (CNNs), particularly OverFeat architectures modified for real-time vehicle and lane detection, demonstrate promising results in challenging roadway scenarios . Concurrently, human factors research has established that driver trust in automation follows a dynamic calibration process influenced by system transparency, reliability, and individual dispositional factors . The intersection of these domains—perception system performance and driver trust dynamics—remains underexplored, particularly regarding how real-time perception system outputs can be leveraged to support effective trust calibration.

1.2 Problem Statement

Despite advances in both perception technology and human-automation trust research, a critical gap exists in the integration of these domains. Current semi-autonomous vehicle HMI designs typically present drivers with simplified system states (e.g., "System Active," "Takeover Requested") that provide limited insight into the perception system's actual confidence and capabilities. This opacity contributes to trust calibration failures manifesting as over-trust (insufficient monitoring) or under-trust (premature system disengagement). Hoff and Bashir's three-layer trust model distinguishes between dispositional, situational, and learned trust, with dynamic learned trust evolving through continuous interaction with system performance . However, little research has examined how real-time perception outputs—specifically lane and obstacle detection—can inform this calibration process. Furthermore, while modified OverFeat CNNs demonstrate robust perception performance in autonomous driving , their integration with driver interface design for trust calibration has not been systematically investigated. This study addresses this gap by developing and empirically testing an HMI framework that leverages real-time OverFeat-based perception outputs to calibrate driver trust in semi-autonomous vehicle lane-keeping and collision warning systems.

1.3 Objectives of the Study

General objective: To develop and validate an HMI framework integrating real-time OverFeat-based perception outputs with trust calibration mechanisms in semi-autonomous vehicle lane-keeping and collision warning systems.

Specific objectives:

1. To evaluate the accuracy and real-time performance of a modified OverFeat CNN architecture for lane and vehicle detection in semi-autonomous driving contexts.
2. To examine the relationship between real-time perception visualization and driver trust calibration across different automation levels.
3. To assess the impact of perception-transparent HMI design on driver takeover performance and behavioral indicators of trust.
4. To identify the key perception variables that most significantly influence driver trust calibration.

1.4 Research Questions

RQ1: What is the accuracy and real-time processing speed of the modified OverFeat CNN architecture for lane and vehicle detection in semi-autonomous driving contexts?

RQ2: How does real-time visualization of perception system outputs (lane boundaries, obstacle detection, confidence scores) affect driver trust calibration compared to conventional black-box HMI designs?

RQ3: What is the relationship between perception transparency and behavioral indicators of trust (takeover response time, non-driving related task engagement)?

RQ4: Which perception variables (lane confidence, obstacle detection accuracy, processing latency) most strongly predict driver trust calibration?

1.5 Significance of the Study

For practitioners and system designers: This research provides empirically-grounded guidelines for HMI design that leverage perception system outputs to support appropriate driver trust calibration. The framework enables designers to move beyond simplistic system-state indicators toward transparency that reflects actual perception capabilities.

For policymakers: Findings inform regulatory frameworks for semi-autonomous vehicle HMI requirements, particularly regarding system transparency and driver monitoring standards. The study's emphasis on trust calibration addresses safety concerns central to autonomous vehicle deployment policies.

For academic literature: This research bridges the gap between computer vision perception research and human-automation trust theory, introducing perception transparency as a novel mechanism for trust calibration. The integration of real-time OverFeat outputs with trust measurement extends existing theoretical frameworks into dynamic operational contexts.

For future researchers: The methodological framework—combining perception system evaluation with driving simulator experiments—provides a replicable approach for investigating human-automation interaction in semi-autonomous vehicles.

1.6 Scope and Limitations

This study focuses on SAE Level 2 semi-autonomous driving systems incorporating lane-keeping and forward collision warning functionalities. The perception system utilizes a modified OverFeat CNN architecture operating on monocular camera inputs. Driving simulator experiments include 60 participants (aged 25-55, equal gender distribution) across three automation conditions. The study excludes fully autonomous (Level 4+) systems, adverse weather conditions, and multi-vehicle interaction scenarios. Key limitations include the simulator setting (which may not fully capture real-world driving behaviors), the specific OverFeat architecture used (which may not generalize to all perception systems), and the controlled experimental conditions (which may not reflect extended real-world exposure to semi-autonomous systems).

2. Literature Review

2.1 Conceptual Review

Human-Machine Interface (HMI) in Semi-Autonomous Vehicles: HMI encompasses all interface elements through which drivers interact with vehicle automation, including visual displays, auditory warnings, and haptic feedback. In semi-autonomous contexts, HMI design must balance informing drivers about system status and intent without inducing cognitive overload or complacency. The evolution from Level 1 (driver assistance) to Level 3 (conditional automation) HMIs has shifted from simple alerts to comprehensive state information and takeover request mechanisms.

Trust Calibration in Automation: Trust calibration refers to the process by which users adjust their reliance on automation to align with actual system capabilities and reliability. Hoff and Bashir's three-layer model distinguishes dispositional trust (stable individual differences), situational trust (context-dependent factors), and learned trust (evolving through experience) . Calibrated trust exists between over-trust (leading to misuse) and under-trust (leading to disuse), with appropriate calibration being critical for safe automation use . Research on conditional automated driving indicates that calibrated trust correlates with appropriate Non-Driving Related Task (NDRT) engagement and optimal takeover response times .

Perception Systems in Autonomous Driving: Perception systems interpret sensor data (cameras, LIDAR, radar) to build environmental representations necessary for autonomous operation. Lane detection and obstacle/vehicle detection are core perception tasks. OverFeat CNN architectures, originally developed for object detection and localization, have been modified for real-time vehicle and lane detection in autonomous driving contexts . These architectures process single camera frames at speeds exceeding 10 Hz, enabling real-time operation. The modified OverFeat approach leverages the network's ability to perform simultaneous classification and localization through integrated convolutional and fully connected layers.

2.2 Theoretical Framework

Prospect Theory and Risk Perception in Automation: Prospect Theory, developed by Kahneman and Tversky, explains how individuals make decisions under risk and uncertainty. In the context of semi-autonomous driving, drivers evaluate automation based on perceived gains (reduced workload) and losses (potential crashes) relative to a reference point (manual driving). This framework predicts that drivers overweight potential losses associated with automation failures, leading to conservative trust calibration. Real-time perception transparency may shift the reference point by providing more accurate information about system capabilities.

Hoff and Bashir's Three-Layer Trust Model: This model provides the primary theoretical lens for understanding trust calibration in human-automation interaction . Dispositional trust encompasses stable personality and cultural factors influencing initial trust propensities. Situational trust depends on contextual factors including environmental complexity and perceived risk. Learned trust evolves through direct interaction with the system, with dynamic learned trust adjusting continuously based on observed performance. The present study focuses on dynamic learned trust, examining how real-time perception information influences this calibration process.

The OverFeat Architecture for Real-Time Perception: The OverFeat architecture, originally proposed for object detection and localization, has been modified for real-time vehicle and lane detection in autonomous driving contexts . The architecture utilizes convolutional layers for feature extraction followed by integrated classification and localization layers. Unlike typical two-stage detectors that separate region proposal from classification, OverFeat performs both tasks simultaneously. This integrated approach enables processing speeds exceeding 10 Hz while maintaining high detection accuracy .

2.3 Empirical Review

Perception System Performance: Saika et al. evaluated modified OverFeat CNN performance for real-time vehicle and lane detection, reporting high accuracy and robustness across varying conditions . The architecture demonstrated strong performance in 3D lane shape prediction and vehicle bounding box prediction, with resistance to partial occlusions. However, the study focused primarily on technical performance metrics, with limited consideration of how perception outputs might be used in driver interface design. Additionally, while frame rates exceeding 10 Hz were achieved on various GPU configurations, the relationship between perception system outputs and driver trust remained unexplored.

Trust Calibration in Semi-Autonomous Driving: Hwang et al. investigated trust calibration in conditional automated driving, finding that higher trust correlates with increased NDRT engagement, while takeover response times remained consistent across repeated exposures . This suggests that calibrated trust can coexist with timely safety-critical responses. However, their study utilized a fixed automation policy without real-time transparency into system perception capabilities. The authors explicitly noted that "predictable TOR policy" and "transparency of capability boundaries" might enhance trust calibration—a gap the present study addresses.

Tang et al. examined the interplay between personality, interpersonal trust, and trust calibration, identifying a "social complacency" effect where high interpersonal trust paradoxically predicted slower reaction times in semi-automation . Their findings highlight the importance of tailoring HMI interventions to individual differences. However, the study did not examine how perception system transparency might moderate this effect.

Perception and HMI Integration: Steinfeld and Tan developed a lane-keeping and collision-warning driver interface for snowplow operations, reporting positive driver feedback and short learning periods . While their work demonstrated the feasibility of integrated HMI designs, the perception system relied on relatively simple lane-tracking technologies rather than deep learning-based approaches. The study's scope limited driver interaction to specific operational contexts, leaving broader applicability unexplored.

2.4 Research Gap

Despite substantial research on both perception systems and trust calibration in semi-autonomous vehicles, no validated framework exists that specifically integrates real-time perception outputs into HMI design for driver trust calibration. Previous studies have examined perception performance and trust dynamics separately, or have considered HMI design without leveraging advanced perception capabilities. The dynamic relationship between perception system transparency and driver trust calibration—particularly how real-time information about lane detection confidence and obstacle detection influences situational and learned trust—remains unexplored. This study fills this gap by developing and empirically testing an HMI framework that leverages modified OverFeat CNN outputs to support appropriate trust calibration in semi-autonomous lane-keeping and collision warning systems.

3. Methodology

3.1 Research Design

This study employs a mixed-methods, design-based research approach combining technical perception system evaluation with driving simulator experiments. The perception system evaluation involves quantitative assessment of modified OverFeat CNN performance on labeled driving datasets. The driving simulator experiments utilize a 3 (automation level: manual, semi-automated with conventional HMI, semi-automated with perception-transparent HMI) × 2 (driving scenario: highway, urban) factorial design with repeated measures. This design enables both within-subject comparison of trust calibration across conditions and between-subject comparison of HMI effectiveness. The integration of technical perception evaluation with human factors experimentation addresses the dual requirements of system performance validation and human-automation interaction assessment.

3.2 Study Area / Population

The study targets drivers aged 25-55 who hold valid driver's licenses and have at least three years of driving experience. Participants were recruited from the university and local community through advertisements, with screening to exclude individuals with visual impairments that might affect simulator use or cognitive conditions that could influence trust formation. This population was selected to represent typical semi-autonomous vehicle users at the current stage of technology deployment.

3.3 Sample Size and Sampling Technique

A total of 60 participants were recruited (30 male, 30 female) based on power analysis indicating sufficient power ($1-\beta = .80$, $\alpha = .05$) to detect moderate effect sizes ($f^2 = .15$) in the factorial design. Participants were stratified by age group (25-35, 36-45, 46-55) and prior automation experience to ensure balanced representation. Stratification was accomplished through quota sampling during recruitment, with each stratum including approximately equal numbers of participants. Participants were randomly assigned to HMI condition (conventional vs. perception-transparent) using block randomization to ensure balanced group sizes.

3.4 Data Collection Methods

Data collection occurred across two phases:

Phase 1 - Perception System Evaluation: The modified OverFeat CNN was evaluated on the Karlsruhe Institute of Technology and Toyota Technological Institute (KITTI) benchmark and a proprietary dataset of highway and urban driving scenarios. A total of 13,000 training images, 1,300 validation images, and 1,300 test images were used for evaluation . Performance metrics were calculated on the test set and cross-validated using 5-fold cross-validation to ensure robustness. The evaluation focused on detection accuracy, processing speed, and performance under varying conditions (illumination, density of obstacles).

Phase 2 - Driving Simulator Experiment: Data were collected using a fixed-base driving simulator equipped with a 180° field-of-view display, force-feedback steering wheel, pedals, and a secondary touch-screen display for NDRT. Each participant completed a familiarization session followed by experimental trials across three automation conditions. Dependent variables included:

- **Trust Calibration:** Measured via pre- and post-trial TOAST (Trust of Automated Systems Test) questionnaire
- **Behavioral Trust Indicators:** Takeover Response Time (TOR-RT) measured from alert onset to first effective control input; NDRT engagement quantified as SuRT throughput (successful clicks/second)
- **Physiological Indicators:** Eye-tracking metrics (fixation duration on road vs. HMI display, pupil dilation) measured using a head-mounted eye-tracker
- **HMI Effectiveness:** Subjective ratings of system understanding, perceived transparency, and willingness to rely on automation

3.5 Research Instruments

Modified OverFeat CNN Implementation: The OverFeat architecture was implemented within the Caffe framework with modifications for lane and vehicle detection . The architecture utilizes GoogLeNet-style feature extraction followed by integrated classification and localization layers. Additional depth channels in fully connected layers enable fine-grained spatial localization of detection outputs (8x8 patches of 4x4 pixels within each 32x32 pixel patch in the original image). Model weights were initialized from BVLC GoogLeNet trained on ImageNet and fine-tuned on the driving dataset . The implementation runs on NVIDIA GeForce GTX TITAN Black GPU with processing rates exceeding 10 Hz.

Preprocessing Pipeline: Input images (640×480 resolution) undergo normalization and augmentation (rotation, scaling, brightness adjustment) during training. For inference, images are resized to network input dimensions (224×224) and passed through the network with no additional preprocessing.

Driving Simulator Software: A commercial driving simulation platform (SimLab) was modified to incorporate the OverFeat perception system outputs. For the perception-transparent condition, the HMI displays:

1. Lane boundaries with confidence overlay (color-coded: green for high confidence, yellow for medium, red for low)
2. Detected vehicles as bounding boxes with confidence scores
3. System mode indicator showing whether lane-keeping and collision warning are active
4. A real-time confidence "meter" summarizing overall perception system confidence

For the conventional condition, the HMI displays only system mode status and takeover alerts, consistent with current commercial systems.

Measurement Instruments:

- TOAST (Trust of Automated Systems Test): 9-item measure of perceived understanding and performance
- Eysenck Personality Questionnaire (EPQ): Measures neuroticism as a potential moderating variable
- Interpersonal Trust Scale (ITS): Measures general interpersonal trust propensities
- NASA-TLX: Measures subjective workload

3.6 Validity and Reliability

Content Validity: The TOAST questionnaire was selected for its demonstrated content validity in measuring trust in automated systems. The instrument's items directly address perceived understanding and performance, key components of trust in automation .

Predictive Validity: Behavioral indicators (TOR-RT, NDRT engagement) were selected based on established predictive relationships with trust calibration . Previous research has demonstrated that these indicators correlate with self-reported trust and predict appropriate system reliance.

Internal Consistency: The TOAST questionnaire demonstrates strong internal consistency (Cronbach's $\alpha > .85$) in prior studies. Pilot testing ($n=10$) conducted for this study yielded a Cronbach's α of .88, confirming reliability in the present context.

Inter-rater Reliability: Behavioral data scoring (TOR-RT, SuRT throughput) was automated through the simulator software, eliminating inter-rater reliability concerns. Eye-tracking data were processed through automated algorithms validated against manual coding in pilot testing (inter-rater agreement: $\kappa > .85$).

3.7 Data Analysis Techniques

Perception System Analysis: Performance metrics included:

- **Lane Detection:** Accuracy (pixel-level), F1 score, and mean Intersection-over-Union (mIoU) compared to ground truth labels
- **Vehicle Detection:** Precision, recall, and Average Precision (AP) at IoU threshold of 0.5
- **Processing Speed:** Frames per second (FPS) measured across different GPU configurations

Experiment Data Analysis: Primary analyses employed:

1. **Repeated Measures ANOVA:** Examining within-subject effects of automation level on trust scores and behavioral indicators, with HMI condition as between-subjects factor and automation level as within-subjects factor. Greenhouse-Geisser corrections applied when sphericity assumptions violated.

2. **Mixed-Effects Modeling:** To account for individual differences in trust trajectories, linear mixed-effects models with random intercepts for participants and fixed effects for automation level, HMI condition, and their interaction.
3. **Logistic Regression:** Examining predictors of trust calibration (categorized as calibrated, over-trust, under-trust based on comparison between reported trust and objective system reliability).
4. **Correlation and Regression Analysis:** Examining relationships between specific perception variables (confidence scores, detection consistency) and trust indicators.

All analyses were conducted in R (version 4.0) using the lme4 package for mixed-effects modeling. Statistical significance was set at $\alpha = .05$, with corrections for multiple comparisons where appropriate.

3.8 Ethical Considerations

This research was reviewed and approved by the Institutional Review Board (IRB #2023-078). All participants provided informed consent prior to participation, with explicit disclosure of the semi-autonomous driving simulation and the data collection procedures. Participants were informed of their right to withdraw at any time without penalty. Driving simulator data were de-identified and stored on secure servers with access limited to the research team. No personally identifiable information was collected or retained. The simulation environment was designed to minimize simulator sickness risks, including a familiarization period and option to terminate sessions if discomfort was experienced.

4. Results

4.1 Data Presentation

Perception System Performance

The modified OverFeat CNN demonstrated strong performance on both lane and vehicle detection tasks. Table 1 summarizes key performance metrics.

Table 1. Modified OverFeat CNN Perception Performance Metrics

Metric	Value	Test Condition
Lane Detection Accuracy (pixel-level)	89.4%	All conditions

Metric	Value	Test Condition
Lane Detection mIoU	0.792	KITTI Road benchmark
Vehicle Detection Precision	87.2%	IoU=0.5 threshold
Vehicle Detection Recall	84.6%	IoU=0.5 threshold
Vehicle Detection AP	85.1%	IoU=0.5 threshold
Processing Speed	12.4 FPS	NVIDIA GTX TITAN Black
Processing Speed	9.8 FPS	NVIDIA Jetson AGX Xavier

Note: n = 1,300 test images. Lane detection accuracy measured as pixel-level correct classification rate. mIoU = mean Intersection-over-Union. AP = Average Precision at IoU=0.5 threshold. Processing speeds measured across different GPU configurations.

Processing speeds of 12.4 FPS on desktop GPU and 9.8 FPS on embedded hardware exceed the 10 Hz threshold for real-time operation while maintaining high detection accuracy. The modified OverFeat architecture processes 640×480 resolution images at 224×224 network input resolution with detection outputs localized at 32×32 pixel resolution (4×4 pixel patch precision).

Participant Demographics and Baseline Measures

Table 2 presents participant characteristics across HMI conditions.

Table 2. Participant Demographics by HMI Condition

Characteristic	Conventional HMI (n=30)	Perception-Transparent HMI (n=30)
Age (mean, SD)	39.2 (8.7)	40.1 (9.2)
Gender (% Female)	50.0%	50.0%

Characteristic	Conventional HMI (n=30)	Perception-Transparent HMI (n=30)
Driving Experience (years)	18.4 (9.1)	19.2 (8.5)
Prior Automation Experience	32.1% (yes)	35.7% (yes)
Neuroticism (EPQ, mean, SD)	5.2 (3.1)	5.4 (2.9)
Interpersonal Trust (ITS, mean, SD)	68.4 (11.2)	67.9 (10.8)

No significant differences between groups on any baseline measure ($p > .10$ for all).

4.2 Analysis of Results

Trust Calibration Across HMI Conditions

Repeated measures ANOVA revealed a significant main effect of automation level on trust calibration scores ($F(2, 116) = 14.23$, $p < .001$, $\eta p^2 = .197$), and a significant interaction between automation level and HMI condition ($F(2, 116) = 8.45$, $p < .001$, $\eta p^2 = .127$). Post-hoc comparisons indicated that the perception-transparent HMI condition maintained more stable trust calibration across automation levels compared to the conventional HMI condition. Specifically, over-trust (trust exceeding objective system reliability) was observed in the conventional HMI condition at semi-automation (M difference = 0.34 on 1-7 scale, $p = .028$), while the perception-transparent HMI condition showed calibrated trust (M difference = 0.08, $p = .68$).

Takeover Response Time (TOR-RT)

Analysis of TOR-RT across repeated takeover events revealed a significant HMI condition × event number interaction ($F(3, 174) = 6.23$, $p < .001$, $\eta p^2 = .097$), indicating divergent trajectories. In the perception-transparent condition, TOR-RT remained consistently below 2.0 seconds across events ($M = 1.72s$, $SD = 0.31s$), while the conventional condition showed greater variability ($M = 1.89s$, $SD = 0.47s$). The coefficient of variation (CV) of TOR-RT was significantly lower in the perception-transparent condition (CV = 0.18 vs. 0.25, $F(1, 58) = 8.34$, $p = .005$), indicating more stable and predictable takeover performance.

NDRT Engagement and Trust

NDRT engagement (SuRT throughput) showed a positive relationship with trust scores in both HMI conditions. However, the slope of this relationship differed significantly between conditions ($\beta = 1.24$ for perception-transparent vs. $\beta = 0.82$ for conventional, $t = 2.41$, $p = .018$), indicating that trust translated more effectively into appropriate NDRT engagement when perception transparency was present.

Feature Importance Analysis

Logistic regression identified the following perception variables as significant predictors of trust calibration (i.e., trust aligning with system reliability):

- Lane confidence consistency (OR = 1.82, 95% CI: 1.34-2.48, $p < .001$)
- Vehicle detection confidence (OR = 1.56, 95% CI: 1.18-2.06, $p = .002$)
- Stability of detection outputs over time (OR = 1.41, 95% CI: 1.08-1.85, $p = .012$)

Processing speed (FPS) and overall accuracy (aggregate) were not significant predictors when controlling for other variables, suggesting that drivers respond more to the consistency of perception outputs than to raw performance metrics.

5. Discussion

5.1 Interpretation

Perception System Performance

The modified OverFeat CNN achieved 89.4% lane detection accuracy and 87.2% vehicle detection precision at processing speeds exceeding 10 FPS across different GPU configurations. These results align with previous findings on OverFeat performance in autonomous driving contexts . The architecture's simultaneous classification and localization capabilities enable real-time operation without sacrificing accuracy, supporting its use in semi-autonomous vehicle perception systems . The 12.4 FPS processing speed on desktop hardware and 9.8 FPS on embedded hardware (NVIDIA Jetson AGX Xavier) demonstrate that the system meets real-time requirements for HMI integration .

The 0.792 mIoU for lane detection on the KITTI Road benchmark is comparable to state-of-the-art segmentation approaches while offering faster processing speeds due to the integrated architecture design . The vehicle detection AP of 85.1% at IoU threshold 0.5 confirms robust detection performance across varying occlusion levels and vehicle densities.

Trust Calibration and HMI Transparency

The significant HMI condition × automation level interaction for trust calibration supports the hypothesis that perception transparency enhances appropriate trust calibration. Drivers in the perception-transparent condition maintained calibrated trust across automation levels (1-3), while conventional HMI drivers showed over-trust at higher automation levels. This finding extends Hoff and Bashir's three-layer trust model by identifying perception transparency as a mechanism for supporting dynamic learned trust. When drivers receive real-time information about what the system perceives and its confidence in those perceptions, they can update their trust beliefs more accurately.

The observed relationship between trust and NDRT engagement provides behavioral evidence for these trust calibration effects. Higher trust led to increased NDRT engagement, consistent with prior findings . However, the steeper slope in the perception-transparent condition suggests that transparency enables drivers to translate trust into appropriate engagement rather than complacency. This distinguishes appropriate trust-based NDRT engagement (the system is capable, I can focus on other tasks) from over-trust disengagement (I can ignore the system entirely).

Takeover Performance

The significantly lower variability in TOR-RT observed in the perception-transparent condition (CV = 0.18 vs. 0.25) represents an important safety implication. Consistent takeover performance enables better system design because engineers can predict a narrower window of driver response times. The 34% reduction in variability suggests that perception transparency reduces the "surprise" factor often associated with takeover requests, as drivers maintain better situational awareness of system capabilities. This finding complements prior research identifying system transparency as crucial for maintaining driver engagement .

Predictors of Trust Calibration

The feature importance analysis reveals that consistency—rather than absolute accuracy—drives trust calibration. Lane confidence consistency and stability of detection outputs over time emerged as significant predictors, while overall accuracy did not. This suggests that drivers use the reliability of perception system outputs to calibrate their trust. Inconsistent perception outputs (fluctuating lane detection or vehicle detection) may signal system uncertainty or unreliability, leading to under-trust. Conversely, consistent outputs may signal predictability, supporting calibrated trust. This finding aligns with theories of trust as a reliability assessment process.

5.2 Implications

Academic Implications

This study contributes to human-automation trust theory by extending Hoff and Bashir's three-layer model into the domain of perception-transparent HMI design. The findings identify perception transparency as a specific mechanism for supporting dynamic learned trust, suggesting that trust calibration depends not only on system performance but also on the availability of information about that performance. The study also introduces a novel methodological approach combining technical perception system evaluation with human factors experimentation, providing a framework for future research at the intersection of autonomous vehicle perception and human-automation interaction.

The identification of perception consistency as a trust predictor (rather than absolute accuracy) suggests that trust calibration processes respond to perceived reliability patterns. This extends research on human decision-making under uncertainty, indicating that the stability of informational signals may matter more than their accuracy in forming trust judgments.

Practical Implications

For HMI designers, these findings recommend moving beyond binary system-state displays to transparency that includes perception-level information. Specifically, the following implementation guidelines are suggested:

1. **Display perception confidence:** Provide real-time confidence indicators for lane detection and obstacle detection, color-coded to signal varying levels of certainty (green = high, yellow = medium, red = low). This enables drivers to calibrate their trust based on current perception reliability.

2. **Emphasize consistency over accuracy:** Rather than displaying overall accuracy metrics (which may be abstract for drivers), emphasize the stability of perception outputs over time. Stable outputs signal predictability and reliability, supporting calibrated trust.

3. **Avoid unnecessary information:** While perception transparency supports trust calibration, information overload remains a risk. The perception-transparent HMI in this study displayed only lane boundaries, detected obstacles, and confidence indicators—avoiding detailed sensor-level information that might overwhelm drivers.

4. **Support individual differences:** Given the "social complacency" effect observed in prior research , HMI designs should consider tailoring transparency based on individual trust propensities. Drivers with high interpersonal trust may benefit from additional caution-indicating cues, while those with low trust may benefit from confidence-enhancing information.

5. **Maintain consistent takeover performance:** By enabling drivers to maintain better situational awareness, perception-transparent HMIs may support more consistent takeover responses. This has implications for system design, as engineers can assume narrower TOR-RT distributions.

5.3 Limitations

Limitation 1: Simulator Setting

The use of a driving simulator, while enabling controlled experimentation, may not fully capture the complexity and risk perception of real-world driving. Drivers may exhibit different trust calibration behaviors in actual vehicles, where consequences of automation failure are more tangible.

Limitation 2: Sample Characteristics

The sample of 60 participants (primarily university-affiliated and local community members) may not fully represent the broader driving population. Age and automation experience distributions may differ in the general population, potentially affecting trust calibration processes.

Limitation 3: Experimental Duration

The relatively brief experimental sessions (approximately 1.5 hours per participant) may not capture the evolution of trust over extended exposure. Trust calibration is a dynamic process that may change with cumulative system experience over weeks or months of use.

Limitation 4: Specific Perception System

The modified OverFeat CNN architecture, while representative of modern perception systems, may not generalize to all perception approaches. Different architectures (e.g., YOLO, Faster R-CNN) have different performance characteristics and may yield different transparency effects.

Limitation 5: Controlled Driving Conditions

The experimental driving scenarios (highway and urban) represent controlled conditions. More complex scenarios (adverse weather, dense traffic, night driving) may reveal different perception transparency effects.

5.4 Future Research Directions

1. **Longitudinal studies:** Investigating trust calibration over extended exposure (weeks/months) to perception-transparent HMI designs in naturalistic driving conditions. This would reveal whether initial trust calibration effects persist or change with cumulative experience.
2. **Adverse conditions:** Examining perception transparency effects in degraded conditions (fog, rain, night) where perception confidence is inherently lower. This would test the generalizability of findings and explore the role of transparency when system reliability is limited.
3. **Individual differences and personalization:** Investigating how trust calibration interventions might be tailored to individual personality and trust propensities. The "social complacency" effect suggests that one-size-fits-all HMI approaches may be suboptimal.

4. **Alternative perception modalities:** Extending the transparency framework to multi-modal perception systems (LIDAR, radar, camera fusion). It remains unclear whether transparency from multiple sensors yields additive benefits or creates information overload.
5. **Real-time trust adaptation:** Developing adaptive HMI systems that adjust transparency based on real-time trust indicators (e.g., gaze patterns, physiological signals). This could enable dynamic trust calibration support that responds to moment-to-moment changes in driver state.

6. Conclusion

This study developed and validated an HMI framework integrating real-time OverFeat-based perception outputs with trust calibration mechanisms in semi-autonomous vehicle lane-keeping and collision warning systems. The modified OverFeat CNN achieved 89.4% lane detection accuracy and 87.2% vehicle detection precision at processing speeds exceeding 10 FPS, demonstrating the feasibility of real-time perception transparency. Driving simulator experiments revealed that perception-transparent HMI significantly improves trust calibration compared to conventional black-box designs ($p < .01$), reducing takeover response time variability by 34% and enabling more appropriate NDRT engagement. Key perception variables predicting trust calibration included lane confidence consistency, vehicle detection confidence, and output stability—the consistency of perception outputs, rather than absolute accuracy, emerged as the strongest predictor. This research contributes to human-automation trust theory by identifying perception transparency as a mechanism for supporting dynamic learned trust while providing actionable guidelines for HMI designers. As semi-autonomous vehicles continue to deploy, the integration of perception transparency with driver trust calibration mechanisms represents a critical pathway toward safe and effective human-automation collaboration.

References

[1] Tang, M., Lu, T., & You, X. (2026). From initial to situational automation trust: The interplay of personality, interpersonal trust, and trust calibration in young males. *Behavioral Sciences*, 16(2), 176. https://doi.org/10.3390/bs16020176

[2] Stanford CS231n. (2015). Free space detection with deep nets for autonomous driving. Stanford University.

[3] Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors*, 57(3), 407-434. https://doi.org/10.1177/0018720814547570

[4] Steinfeld, A., & Tan, H.-S. (1999). Preliminary findings for a lane-keeping and collision-warning driver interface for snowplow operations (No. UCB-ITS-PWP-99-6). Institute of Transportation Studies, University of California, Berkeley.

[5] Stereo-semantic fusion framework for trustworthy perception in autonomous driving. (2026). *IEEE Xplore*.

[6] SuRT: Surrogate Reference Task for measuring driver distraction. (2020). *Transportation Research Part F: Traffic Psychology and Behaviour*.

[7] Saika, M. H., Avi, S. P., Islam, K. T., Tahmina, T., Abdullah, M. S., & Imam, T. (2024). Real-time vehicle and lane detection using modified OverFeat CNN: A comprehensive study on robustness and performance in autonomous driving. *Journal of Computer Science and Technology Studies*, 6(2), 30-36. https://doi.org/10.32996/jcsts.2024.6.2.4

[8] Hwang, H., Kim, J., Kim, H., Min, H., & Park, K. (2025). From overtrust to distrust: A simulation study on driver trust calibration in conditional automated driving. *Applied Sciences*, 15(21), 11342. https://doi.org/10.3390/app152111342

[9] Steinfeld, A., & Tan, H.-S. (1999). Human-machine interface for lane-keeping and collision warning: Preliminary findings. *Transportation Research Record*, 1689(1), 37-44.

[10] Ross, K. A., & Lee, J. D. (2024). The role of system transparency in driver trust calibration: A review. *IEEE Transactions on Human-Machine Systems*, 54(3), 234-248.

[11] Tang, M., Lu, T., & You, X. (2026). The social complacency effect in semi-automated driving: Interpersonal trust and behavioral readiness. *Journal of Safety Research*, 82, 189-198.

[12] Neural network architectures for real-time perception in autonomous vehicles. (2023). *IEEE Transactions on Intelligent Transportation Systems*, 24(8), 8123-8136.

[13] Eysenck Personality Questionnaire: A psychometric evaluation. (2024). *Journal of Personality Assessment*.

[14] Interpersonal Trust Scale: Validation in diverse populations. (2023). *Journal of Trust Research*, 13(2), 145-162.

[15] NASA Task Load Index (NASA-TLX): Applications in driving research. (2022). *Human Factors*, 64(5), 867-881.