Academic Journal of Engineering and Technology Science
ISSN 2616-5767 Vol.8, Issue 4: 67-71, DOI: 10.25236/AJETS.2025.080409

Design of Human-Machine Interaction for Autonomous Vehicles in the Era of Intelligence

Hao Tang

Sichuan Yiyun Intelligent Connected Vehicle Technology Co., Ltd., Yibin, Sichuan, China

Abstract: *Against the backdrop of rapid development in artificial intelligence and sensor technology, autonomous vehicles are steadily moving toward practical application. An efficient, natural, and safe human-machine interaction (HMI) system has become the core for gaining user trust and achieving wide-scale adoption. Focusing on the specific scenario of L4-level autonomous driving, this study proposes an HMI framework integrating multimodal perception, adaptive decision-making, and cognitive load optimization. It particularly examines deep learning-based intent recognition algorithms, dynamic optimization strategies for takeover requests in emergency situations, and multi-channel fusion mechanisms for presenting interactive information. Through high-fidelity driving simulators and real-vehicle road tests, combined with eye-tracking, physiological signal monitoring, and other relevant methods, the system's effectiveness in reducing cognitive load, enhancing situational awareness, and shortening takeover reaction time was validated. Data show that the optimized HMI reduces average driver takeover reaction time by 42.7%, increases situational understanding accuracy to 91.3%, and significantly lowers subjective workload index. This research provides key technical support for building a highly usable and safe human-machine co-driving paradigm for autonomous vehicles and holds great significance for promoting the commercialization of autonomous driving.*

Keywords: *Autonomous Vehicles; Human-Machine Interaction Design; Multimodal Interaction; Cognitive Load; Takeover Performance; Reinforcement Learning; Eye-Tracking*

1. Introduction

The automotive industry is undergoing profound changes centered on electrification, intelligence, and connectivity, with high-level autonomous driving technology becoming the focus of global competition. According to SAE's definition, autonomous driving systems at Level 3 and above are required to perform all dynamic driving tasks within specific operational design domains (ODD), but still need human takeover when the system fails or exceeds the ODD. This role shift presents serious human factors engineering challenges. In autonomous driving mode, drivers are in an "out-of-the-loop" state, with significantly diminished situational awareness. When faced with sudden system takeover requests, they often exhibit delayed reactions and decision-making errors [1]. Traditional in-vehicle HMI design paradigms can no longer meet the complex demands of situational awareness reconstruction, trust building, and efficient collaboration in autonomous driving scenarios. Therefore, designing intelligent and adaptive HMI systems to achieve human-vehicle collaborative decision-making and smooth control authority transfer has become a core issue for ensuring the safety and user experience of autonomous driving.

2. HMI System Architecture and Interaction Framework Design for Autonomous Driving

The HMI system for L4-level autonomous driving needs to establish a multi-level perception–decision–execution closed loop, with its core architecture consisting of the environmental perception layer, user state understanding layer, interaction decision-making layer, and multimodal execution layer. The environmental perception layer integrates LiDAR point clouds, camera images, millimeter-wave radar data, and V2X information to construct a 360-degree dynamic environmental model around the vehicle. The user state understanding layer incorporates driver-facing DMS cameras, steering wheel grip sensors, microphone arrays, and bioelectrodes to capture in real time the driver's eye movement trajectory, head posture, voice commands, and physiological signals [2].

The interaction decision-making layer, as the core of the system, adopts a hierarchical reinforcement learning framework. The high-level policy determines the current interaction mode based on the environmental risk level and the driver's cognitive state, while the low-level policy dynamically generates specific interaction content and timing. The system's state space S includes environmental risk factors $E_r \in R^5$(such as time to collision, TTC, road curvature, weather visibility, etc.), driver state factors $D_s \in R^4$ (such as attention dispersion index, physiological arousal level, fatigue level), and vehicle state $V_s \in R^3$(speed, acceleration, steering angle). The action space A includes interaction modality selection, information abstraction level, and presentation timing.The reward function is designed as:

$$R(s,a)=w_1 \cdot SA(s')+w_2 \cdot (1-CL(s'))+w_3 \cdot TTR(s')-w_4 \cdot E_c \tag{1}$$

In Equation 1, SArepresents situational awareness, reflecting the driver's understanding of the current driving environment and vehicle state; CL represents cognitive load, indicating the mental burden on the driver when processing information; TTR represents the predicted takeover reaction time, i.e., the estimated time from when the driver receives a takeover request to actually taking control of the vehicle;$E_c$represents interaction energy consumption, referring to the energy consumed by the system when performing interaction operations. The weight coefficients are optimized through inverse reinforcement learning to ensure that the reward function accurately reflects the system's prioritization of different indicators. (Table 1)

Table 1 Correspondence between HMI Interaction Modes and Trigger Conditions

Environmental Risk Level	Driver State	Preferred Interaction Mode	Information Abstraction Level	Haptic Feedback Intensity
Low risk	High alertness / active monitoring	Visual summary + auditory status prompt	Low (icon-based)	Weak vibration
Medium risk	Moderate attention / mild distraction	Highlighted visual prompt + voice alert	Medium (semantic brief)	Moderate pulse
High risk	Low alertness / deep distraction	Full-modal alert + steering wheel vibration	High (detailed instructions)	Strong continuous vibration
Emergency takeover request	Any state	Multi-channel redundant alert + forced intervention	Highest (action-oriented)	Maximum intensity

3. Multimodal Perception and Intention Understanding Technology

Accurate recognition of driver intention is a prerequisite for achieving proactive interaction. In this study, a driver state recognition model is constructed based on the fusion of multi-source heterogeneous data. Eye tracking employs an infrared camera with a sampling rate of 250 Hz to extract features such as fixation point sequence, saccade velocity, and pupil diameter change rate [3]. A spatio-temporal graph convolutional network (ST-GCN) is used to model the dynamics of eye behavior:

$$F_{gaze}=\sigma\left(W*concat[X_t,X_{t-\Delta t},\cdots,X_{t-n\Delta t}]+b\right) \tag{2}$$

In Equation 2, X_t represents the eye movement feature vector at time t, W denotes the convolution kernel weights,b is the bias term, and σ is the activation function. This formula is used to obtain the eye movement features F_{gaze}.

Physiological signals collected include EEG, ECG, and skin conductance response. EEG focuses on the frontal lobe θ wave to parietal lobe α wave power ratio to assess cognitive load; ECG extracts the RMSSD indicator of heart rate variability to reflect psychological stress. A multimodal Transformer fusion model is employed:

$$Attention(Q,K,V)=softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{3}$$

$$Z=LayerNorm(FFN(Attention(F_{gaze},F_{bio},F_{posture}))) \tag{4}$$

In Equations 3 and 4,Q , K, and V represent the query, key, and value matrices, respectively; d_k is the dimension of the key vector; FFN denotes the feedforward neural network; LayerNorm represents the layer normalization operation. The output driver state $D_s \in R^4$vector includes attention level, cognitive load index, emotional valence, and fatigue level.

To improve the model's accuracy and robustness, denoising, normalization, and other preprocessing operations are performed on eye movement data and physiological signals during the data preprocessing stage. For outliers in eye movement data, statistical-based methods are used for detection and correction; for baseline drift in physiological signals, filtering algorithms are applied to eliminate it. During model training, cross-validation is employed to ensure the model maintains good performance across different datasets. Verified by extensive experimental data, this multimodal fusion model achieves over 90% accuracy in recognizing driver states, providing a reliable basis for subsequent interaction decision-making.

4. Cognitive Load Optimization and Information Presentation Strategy

To avoid information overload, a cognitive load quantification model needs to be established. Dual calibration is performed using the NASA-TLX subjective evaluation and the objective indicator of pupil diameter change rate. The cognitive load index is defined as:

$$\mathrm{CLI}=\alpha\cdot\frac{\Delta \mathrm{PD}}{\mathrm{PD}_{\mathrm{baseline}}}+\beta\cdot\sum_{i=1}^{3} w_i\,\mathrm{TLX}_i \tag{5}$$

In Equation 5,$\Delta \mathrm{PD}$ represents the standard deviation of pupil diameter changes; $\mathrm{PD}_{\mathrm{baseline}}$ is the baseline pupil diameter; TLX_i are the scores for the mental demand, temporal demand, and effort dimensions respectively; $\alpha$and $\beta$are weighting coefficients;$w_i$denotes the weights of each TLX dimension, calibrated through experimental data.

Based on the CLI, the information flow is dynamically adjusted: when CLI > 0.65, an information filtering mechanism is activated, presenting only critical navigation and safety information to reduce unnecessary distractions for the driver; when 0.4 < CLI ≤ 0.65, key routes are projected via AR-HUD, allowing the driver to access driving-related information more intuitively; when CLI ≤ 0.4, information about surrounding points of interest can be provided to enrich the driver's travel experience. Visual coding follows the ISO 2575 standard, using color semantics: red indicates immediate action, yellow indicates warning, and green indicates normal status (Table 2).

Table 2 Multi-Channel Information Encoding Specifications

Information Type	Visual Channel	Auditory Channel	Haptic Channel	Presentation Duration
Navigation Instructions	Arrow + distance overlay	3-tone prompt sound	None or slight single vibration	≤ 2 seconds
Forward Collision Warning	Red flashing frame	Urgent beep (850 Hz)	High-intensity continuous vibration	Until danger is cleared
Lane Departure	Yellow border flashing	Medium-frequency prompt (600 Hz)	Single-side seat vibration	≤ 3 seconds
System Mode Switch	Status bar color gradient	Voice announcement "Autonomous Driving Activated"	Double pulse vibration	1.5 seconds

In the timing of information presentation, priority is given based on the urgency and importance of the information. For urgent and important information, such as forward collision warnings, presentation should be prioritized; for non-urgent but important information, such as navigation instructions, presentation can occur at appropriate moments. Meanwhile, considering the driver's visual attention allocation, sudden presentation of large amounts of information should be avoided when the driver's gaze is focused on the road ahead to prevent impacting driving safety. Using eye-tracking technology to monitor the driver's fixation points in real time, the information presentation time is appropriately extended when the driver's gaze is on key information areas to ensure accurate information acquisition.

5. Safety Takeover and Collaborative Decision-Making Mechanism

The takeover process is a core challenge in autonomous driving HMI design. A takeover timing optimization model based on the Markov Decision Process is proposed. Takeover readiness is defined as:

$$\mathrm{R}_{\mathrm{takeover}}=\frac{1}{1+e^{-(k_1\cdot \mathrm{TTC}^{-1}+k_2\cdot D_{\mathrm{aware}}-k_3\cdot C_{\mathrm{complex}})}} \tag{6}$$

In Equation 6,TTC represents time to collision,D_{aware} is the driver's situational awareness score, C_{complex} denotes road complexity, k_1、$k_2$and $k_3$is a model parameter obtained through experimental data fitting.

Takeover guidance strategy is established as follows: When $\mathrm{R}_{\mathrm{takeover}}>0.8$ occurs, an emergency takeover protocol is triggered, during which the system issues a strong takeover request to the driver through multiple channels and prepares for forced intervention; when $0.6<\mathrm{R}_{\mathrm{takeover}}\leq 0.8$ occurs, a level-3 progressive warning is initiated, gradually increasing the intensity of information prompts to guide the driver to prepare for takeover; when $\mathrm{R}_{\mathrm{takeover}}\leq 0.6$ occurs, only a status prompt is given to inform the driver of the current vehicle and environmental conditions. Dynamic time window adjustment is adopted:

$$T_{lead}=\max\left(T_{min},\frac{T_{base}}{1+\gamma\cdot(1-D_{aware})}\right) \quad (7)$$

In Equation 7, T_{base} is the baseline guidance time (experimentally calibrated as 6.3 seconds), T_{min} is the minimum guidance time, γ is the awareness decay coefficient, and D_{aware} is the driver's situational awareness score.

At the collaborative decision-making layer, a hybrid reinforcement intelligence framework is designed. When the system confidence exceeds 95%, decisions are executed automatically without driver intervention; when confidence is between 80% and 95%, suggested plans are proposed for driver confirmation, allowing the driver to decide whether to adopt them based on their judgment; when confidence is below 80%, control is handed over and manual decision-making is prompted to ensure driving safety. The confidence calculation model is:

$$C_{sys}=1-\prod_{i=1}^{N}(1-p_i\cdot w_i) \quad (8)$$

In Equation 8, p_i represents the confidence of each perception module; w_i is the dynamic weight, adjusted according to the performance of each perception module and the complexity of the current environment; N denotes the number of perception modules.

To verify the effectiveness of the safety takeover and collaborative decision-making mechanisms, extensive simulation experiments and real-vehicle tests were conducted. In the simulation experiments, various complex traffic scenarios such as sudden obstacles and adverse weather conditions were set up to test the system's takeover performance and collaborative decision-making effectiveness under different situations. Experimental results show that the mechanism can quickly and accurately make decisions in various scenarios, effectively shortening takeover reaction time and improving driving safety. In the real-vehicle tests, multiple drivers were invited to participate, and their operational data and subjective evaluations in different takeover scenarios were collected, further validating the mechanism's practicality and reliability [4].

6. Experimental Validation and Performance Analysis

A testing environment was established based on the CARLA simulation platform and a real-vehicle modification platform. The CARLA simulation platform can simulate various complex traffic scenarios and road conditions, providing rich test cases for the experiments; the real-vehicle modification platform conducts tests in real road environments to ensure the authenticity and reliability of the experimental results. Thirty-two participants (equal numbers of males and females, with driving experience over three years) were recruited to test takeover performance in simulated urban road scenarios. Six typical takeover scenarios were designed: sudden braking of the vehicle ahead, pedestrian crossing, lane change in construction zones, system failure, adverse weather, and V2X emergency messages.

Table 3 Comparison of Key Indicators before and after HMI Optimization (n=32)

Performance Indicator	Traditional HMI	This Design HMI	Improvement	Significance (p-value)
Average Takeover Reaction Time (ms)	3,812	2,183	-42.7%	<0.001
Takeover Operation Accuracy	76.4%	93.8%	+17.4%	0.002
Situational Understanding Accuracy	72.1%	91.3%	+19.2%	<0.001
NASA-TLX Total Score	68.3	42.7	-37.5%	0.001
User Satisfaction Score	5.8/10	8.7/10	+50.0%	<0.001

As shown in Table 3, physiological data analysis shows that the optimized HMI reduced drivers' peak skin conductance response during takeover by 31.2%, indicating a significant decrease in stress response; heart rate variability increased by 27.5%, demonstrating a more stable psychological state. Eye-tracking heatmap analysis confirms that the new design increased fixation concentration on key information areas to 89%, improving visual search efficiency by approximately 40%, enabling drivers to acquire critical information faster and more accurately.

Data collected from real-vehicle road tests in high-speed scenarios show that at a speed of 80 km/h, the system can complete the entire process from driver state assessment to control handover within 3.2 seconds, meeting the time constraints for emergency operations specified by ISO 26262. In-depth analysis of the experimental data reveals that the designed HMI outperforms traditional HMI across different scenarios. In emergency situations, such as sudden braking of the vehicle ahead and pedestrian crossing, the reduction in average takeover reaction time is more pronounced, which is crucial for preventing traffic accidents. Meanwhile, drivers reported high satisfaction with the designed HMI, considering it easy to operate and clear in information presentation, effectively reducing driving workload.

7. Conclusion

This study addresses the human-machine interaction challenges of L4-level autonomous vehicles by proposing an advanced HMI architecture that integrates multimodal perception, cognitive load optimization, and reinforcement learning-based decision-making. By constructing a driver state recognition Transformer model, a dynamic takeover readiness evaluation algorithm, and multi-channel information encoding specifications, the human-machine collaboration efficiency is significantly improved. Experimental data validate the superiority of this design in shortening takeover reaction time, enhancing situational understanding accuracy, and reducing cognitive load. The research outcomes provide key technical support to overcome the bottlenecks in autonomous driving deployment, with future work aiming to deepen interaction allocation among multiple passengers. This technical framework has applied for multiple patents and completed testing on prototype vehicles for automotive manufacturers, establishing a new benchmark in human factors engineering design for intelligent driving and bearing great significance for promoting the commercialization of autonomous driving.

References

[1] Wang Hanjie. Design of Autonomous Vehicle Onboard Interaction Interface Based on Trust Structure Theory [J]. Silk Screen Printing, 2024,(21):98-100.

[2] Wang Yuyang. Research on Intelligent Human-Machine Interaction System for Autonomous Driving Based on Deep Learning [D]. Chang'an University, 2024.

[3] Zhang Hui, Nie Zhiling, Xiao Hongwei, et al. Color Design of Human-Machine Interaction Interface in Intelligent Cockpit of Autonomous Vehicles [J]. Journal of Jilin University (Engineering Edition), 2023, 53(05):1315-1321.

[4] Chen Liwei. Human-Machine Interaction Design for Autonomous Vehicles in the Intelligent Era [D]. Southeast University, 2022.