A Multi-Agent AI-Driven Framework for Driver State Estimation and Intervention: A Modular Approach to Enhance Takeover Performance

Yubo Jiao[a], Luis Miranda-Moreno[a], and Jiangbo Yu[a, *]

[a] *Department of Civil Engineering, McGill University, Montreal H3A 0G4, Canada*

ABSTRACT
Automated driving mitigates the adverse effects of drivers' negative mental and physiological states, while concurrently posing novel challenges in sustaining their optimal well-being. Drivers must assume manual control when the automated system reaches its operational limits, with takeover performance strongly influenced by the driver states. To improve drivers' takeover performance, we propose a framework that employs multiple artificial intelligence (AI)-based agents for real-time estimation and intervention of driver states. Within the proposed framework, we model dynamic driver states base on state-space representation to architect dual AI-based agents: a convolutional neural network (CNN)-based state estimator and a large language model (LLM)-driven intervenor. The estimation agent employs hybrid attention mechanisms to decode temporal dependencies in multimodal physiological and biological signals and vehicular parameters. Concurrently, the intervention agent assesses state-context risk and provides context-aware mitigation strategies to optimize human-machine coordination. We demonstrated the efficacy of our framework through a case study on driver distraction. Driver distraction was estimated in real-time by analyzing multiple indicators, including driving speed, duration, eye movements, self-reported feelings, blood volume pulse, and electrodermal activity. Subsequently, we employed an LLM-powered chatbot for context-sensitive persuasion strategies implementation to alert drivers. Our findings demonstrate that the proposed framework provides an effective state-space representation for understanding and modeling of driver states. The framework implements modular AI components to employ multiple techniques for fulfilling distinct functional objectives. Moreover, the framework's architecture facilitates extensibility beyond driver state management via its modular design, providing a generalizable formalism for human-machine system governance.

Keywords: Driver state, state estimation, state intervention, multi-agent systems, large language models (LLMs), artificial intelligence

INTRODUCTION
Drivers are the central component of the human-vehicle system, making their mental and physiological states critical for ensuring the system's overall safety and efficiency. Adverse driver states compromise situational awareness to diminish responsiveness to the driving environment, delaying emergency reactions and elevating the risk of accidents (*1*). A Transportation Canada report indicates that impaired driving states consistently contributed to fatal collisions between 2018 and 2022 (*2*). Autonomous vehicles can enhance traffic safety by supporting functions such as lane-keeping and car-following; however, drivers must be prepared to retake control when the system issues a takeover request—for example, in critical situations or on road segments beyond the system's operational capabilities (*3*). Thus, timely assessing driver state and implementing appropriate interventions are essential for enhancing driver takeover performance.

Driver states are latent varaibles that cannot be measured directly; however, variations in observable physiological and biological signals can serve as proxies for driver states (*4*). These latent states exhibit temporal dependencies, wherein the current state is influenced by both preceding states and prevailing environmental conditions. State-space models offer an effective solution for modeling dynamic driver states (*5*). Using such state-space representation, we can characterize the evolution of driver states by incorporating historical data and driving contexts. This characterization enhances our understanding of how driving contexts influence driver states. Furthermore, such state-space representation elucidates the effect of these states on observable driver responses, providing a theoretical basis for inferring latent states from observations. Therefore, modeling dynamic driver states using a state-space representation deepens our understanding of driver state estimation and informs effective intervention strategies.

Artificial intelligence (AI) offers the potential for real-time estimation and intervention in driver states. Specifically, machine learning techniques are extensively employed to develop mappings that correlate latent driver states with both driving contexts and observable driver responses. In particular, convolutional neural networks (CNNs) can capture the nonlinear relationships between driver states and their responses, thereby enabling more reliable and accurate state evaluations. Furthermore, large language models (LLMs) can interpret driving contexts in a human-like manner, allowing them to engage with drivers and provide tailored feedback (*6*). This capability facilitates the customization of intervention strategies in real time. By leveraging multimodal large language technologies that integrate visual and auditory modalities, these intervention strategies can be effectively implemented. Therefore, this study aims to investigate the application of convolutional neural networks for driver state estimation and LLM-based voice chatbots for driver state intervention.

MATERIALS AND METHODS

The driver state model describes the evolution of the driver's state, $\boldsymbol{x}_t$, over time. At time t, the driver state is updated based on its current state and the driving contexts, $\boldsymbol{u}_t$. In this study, we assume the driver state system to be a nonlinear dynamical system. Therefore, the driver state model based on a nonlinear state-space representation can be represented as follows:

$$\dot{\boldsymbol{x}}_t = \boldsymbol{f}(\boldsymbol{x}_t, \boldsymbol{u}_t) \tag{1}$$

$$\boldsymbol{y}_t = \boldsymbol{h}(\boldsymbol{x}_t, \boldsymbol{u}_t, \boldsymbol{n}_t) \tag{2}$$

where $\boldsymbol{y}_t$ represents the observations of driver's responses at time t, $\boldsymbol{f}(\cdot)$ is the nonlinear vector function defining the dynamics of the driver states, $\boldsymbol{h}(\cdot)$ is the nonlinear vector function mapping the driver states and driving contexts to driver responses, and $\boldsymbol{n}_t$ represents the observation noise.

We introduce an estimation agent based on CNNs for estimating the latent state, $\boldsymbol{x}_t$, from observed driver responses, $\boldsymbol{y}_t = \left\{\boldsymbol{y}_t^{(1)}, \boldsymbol{y}_t^{(2)}, \cdots, \boldsymbol{y}_t^{(p)}\right\} \in \mathbb{R}^p$. In particular, we employ a multi-input CNNs architecture to learn a mapping between multiple responses and driver states. This estimation agent utilizes CNNs for feature extraction from each response to handle multiple modalities and the inherent nonlinearities. The estimation agent approximates the posterior distribution to evaluate the driver states:

$$\widehat{\boldsymbol{x}}_t = \arg\max_K P(\boldsymbol{x}_t = K \mid \boldsymbol{y}_t; \Theta) \tag{3}$$

Where K is the one of driver states, and Θ encapsulating all learnable parameters of the network.

Furthermore, we introduce an LLM-based intervention agent that modulates driving contexts, $\boldsymbol{u}_t$, to influence latent states, $\boldsymbol{x}_t$. Specifically, the agent employs an LLM to assess the risk associated with driver states within given driving contexts and to generate context-aware mitigation strategies that optimize these latent states. This intervention is implemented via multimodal LLMs, which deliver strategies through visual and auditory modalities. The agent incorporates feedback control with prior knowledge to effectuate driver state intervention:

$$\boldsymbol{u}_t = \boldsymbol{\mathcal{K}}\left(\boldsymbol{x}_t^{ref} - \boldsymbol{x}_t\right) \tag{4}$$

Where $\boldsymbol{\mathcal{K}}(\cdot)$ is the feedback function determined by LLMs, and $\boldsymbol{x}_t^{ref}$ represents the desired target states.

RESULTS AND DISCUSSION

We demonstrated the efficacy of our framework through a case study on driver distraction. Real-time estimation of distraction was achieved by analyzing multiple indicators—including driving speed, duration, eye movements, self-reported feelings, blood volume pulse, and electrodermal activity—which confirmed that the estimation agent reliably detects driver distraction. Furthermore, an LLM-powered chatbot was employed to implement context-sensitive persuasion strategies for alerting drivers. Experimental results indicate that the chatbot effectively engages drivers and provides timely reminders.

CONCLUSION

This study proposes a framework that integrates multiple AI-based agents to address driver state estimation and intervention using a dynamic state-space representation. Within this framework, we introduce two real-time agents: (1) an estimation agent that leverages CNNs to infer latent driver states from temporal dependencies in multimodal physiological and biological signals, as well as vehicular parameters; and (2) an intervention agent that employs LLMs to assess state-context risk and deliver context-aware mitigation strategies for optimizing human–machine coordination. This framework advances our understanding of dynamic driver states by enabling modular AI components tailored to distinct modeling objectives. Consequently, it facilitates timely estimation and intervention, thereby enhancing human–vehicle interaction and overall driving safety.

REFERENCES

[1] Dong, Y. C., Z. C. Hu, K. Uchimura, and N. Murayama. Driver Inattention Monitoring System for Intelligent Vehicles: A Review. *Ieee Transactions on Intelligent Transportation Systems,* Vol. 12, No. 2, 2011, pp. 596-614.

[2] Transport Canada. *Canadian Motor Vehicle Traffic Collision Statistics: 2022*. https://tc.canada.ca/en/road-transportation/statistics-data/canadian-motor-vehicle-traffic-collision-statistics-2022.

[3] Gasne, C., L. Paire-Ficout, S. Bordel, S. Lafont, and M. Ranchet. Takeover performance of older drivers in automated driving: A review. *Transportation Research Part F-Traffic Psychology and Behaviour,* Vol. 87, 2022, pp. 347-364.

[4] Tavakoli, A., S. Boker, and A. Heydarian. Driver State Modeling Through Latent Variable State Space Framework in the Wild. *Ieee Transactions on Intelligent Transportation Systems,* Vol. 24, No. 2, 2023, pp. 1879-1893.

[5] Hamilton, J. D. State-space models. *Handbook of econometrics,* Vol. 4, 1994, pp. 3039-3080.

[6] Cui, C., Y. Ma, X. Cao, W. Ye, and Z. Wang. Drive as you speak: Enabling human-like interaction with large language models in autonomous vehicles.In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2024. pp. 902-909.