# From Gaze to Movement: Predicting Visual Attention for Autonomous Driving Human-Machine Interaction based on Programmatic Imitation Learning (Supplementary Material)

Anonymous ICCV submission

Paper ID 14229

This document provides further details on our study, including dataset specifications and experimental methodologies. Additional information on DATAD is available in Section S-1.

## S-1. DATAD

In normal driving scenarios, drivers may experience lapses in visual attention, but the safety impact is significantly lower compared to high-risk scenarios, such as rear-end collisions, where the consequences of errors are far more severe. However, since such critical incidents are relatively rare, collecting sufficient driver eye movement data in these scenarios is challenging. To address this issue, prior research has explored the use of simulated driving tasks in laboratory environments to collect eye movement data. The BDD-A experimental protocol has demonstrated that laboratory-collected visual attention data can reliably identify the critical visual regions drivers focus on in potentially hazardous situations[7].

We present DATAD, a newly constructed dataset for autonomous driving takeovers. DATAD includes 12 types of takeover scenarios, comprising approximately 600,000 frames of driving data, integrating both driver operation data and surrounding traffic flow information. This dataset provides a richer resource for studying visual attention in human-machine interaction within autonomous driving contexts.This document provides further details on our study, including dataset specifications and experimental methodologies. Additional information on DATAD is available in the this section.

### S-1.1. The Data Sources of the DATAD Dataset

DATAD is a driver takeover eye-tracking dataset designed to meet the emerging human-machine interaction research needs in autonomous driving development. Building upon the standardized experimental design protocols established by BDD-A[7] and DADA[2], we introduce a high-fidelity driving simulator with a real driver, traffic flow, and an in-the-loop real vehicle to capture eye-tracking data during actual driving scenarios.

| Gender | Age | Have a driver's license | Experience in autonomous driving |
|---|---|---|---|
| 22 female, 38 male | 21-48 | 100% | 100% |

Table S-1. Main Statistical Information of Participants Collected from the Questionnaire

**Participants** This study recruited 60 participants who met the experimental requirements, including 22 females (36.7%) and 38 males (63.3%). All participants held a valid driver's license, were aged between 21 and 48, had normal or corrected-to-normal vision, and possessed basic proficiency in using smart devices. To ensure sample diversity, participants were recruited through online registration, campus announcements, and social media platforms. Strict selection criteria were applied to exclude individuals who had experienced severe traffic accidents in the past year, were taking medication that could affect cognition, or suffered from severe motion sickness. Additionally, participants with prior experience in similar research were excluded to minimize learning effects. Before the experiment, all participants provided informed consent and completed a demographic questionnaire covering gender, age and autonomous driving experience. Detailed participant data can be found in Table Tab. S-1

**Experimental Equipment:** As shown in Figure S-1, our full-scale driving simulator allows a real vehicle to be positioned directly onto the experimental platform, enabling participants to conduct driving tasks within an actual vehicle cabin. The simulation environment features a curved projection screen and immersive display system, providing a highly realistic visual experience. Steering, braking, and acceleration inputs are seamlessly synchronized between vehicle sensors and simulation software, ensuring that drivers receive authentic control feedback similar to real-world driving conditions. Additionally, the dynamically responsive rig, powered by a synchronized belt-driven dynamometer, delivers precise directional control, replicating real vehicle dynamics. A vehicle domain controller further integrates physical signals and virtual environments, allowing for realistic interaction in complex simulation scenarios—making it an advanced testbed for human-machine interaction research in autonomous driving. Furthermore, the eye-tracking system utilizes a Smart-eye remote eye tracker, which continuously monitors driver gaze trajectories, fixation distribution, and attention shifts in real time. This system enables the analysis of cognitive load, visual attention allocation, and potential fatigue across different driving scenarios.

Figure S-1. Experimental Equipment: High-fidelity driving simulator (left) and in-vehicle eye-tracking system (right)

### S-1.2. Experimental Procedure

Participants performed takeover operations in 12 autonomous driving takeover scenarios, each structured into four phases: autonomous driving phase → takeover request (TOR) issued → takeover operation phase → scenario completion. The autonomous driving phase lasted approximately 40 seconds, during which participants were required to look at their smartphones, simulating a non-driving state[1, 4, 6]. Upon reaching the takeover trigger point, the system issued a verbal TOR prompt instructing the driver to take control. After hearing the TOR, the driver had to press a button to switch from autonomous to manual mode, then visually assess the environment and execute the appropriate takeover action. Each scenario had a designated takeover endpoint, marking the completion of the trial. Since some takeover failures (e.g., collisions with other vehicles or infrastructure) could occur, not all drivers successfully reached the endpoint. However, regardless of success or failure, the eye movement transition from distraction to active driving state was recorded, making all trials valuable for analysis. To minimize order effects, we used a Latin square design to randomize the sequence of the 12 scenarios, with each participant blindly selecting a scenario order[3]. Each trial lasted approximately 2 ± 1 minutes, with a total experimental duration of around 30 minutes per participant. No participants reported fatigue or motion sickness, and the study received ethical approval from the university's Institutional Review Board (IRB). During data collection, participants first entered the real vehicle cabin and adjusted the seat position for comfort and natural driving posture. The steering wheel, accelerator, and brake pedal sensors were then tested to ensure accurate response and realistic control feedback. Once the driving posture was stabilized, researchers calibrated the eye-tracking system to ensure data accuracy and reliability. Finally, participants underwent a 5-minute adaptation phase, freely driving in CARLA's default urban environment to familiarize themselves with the simulator, enhance immersion, and replicate real-world driving conditions as closely as possible.

### S-1.3. Takeover Scenario Design

In this study, we systematically analyzed takeover-related literature, scenario videos, and accident reports to extract key abstract elements such as vehicle motion trajectories, traffic participants, and scene characteristics[5]. The scenario design process consisted of the following steps:(1)Scenario Selection and Construction: We identified representative emergency takeover cases from literature, videos, and reports, integrating common real-world traffic situations, such as congested urban intersections and main roads with side road entries, to construct realistic experimental scenarios. (2)Dynamic Element Simulation: Each scenario incorporated dynamic traffic objects to simulate realistic traffic flow and unexpected events, such as sudden braking of leading vehicles or pedestrians unexpectedly crossing the road. (3)Takeover Parameter Configuration: Each scenario included a 40-second non-driving task to ensure participants were in a distracted state before the takeover event. Additionally, the takeover completion point remained consistent across scenarios to maintain spatial uniformity in the takeover process. (4)Following this process, we designed 12 representative emergency takeover scenarios. These scenarios are not only highly realistic but also comprehensively assess drivers' response capabilities in critical situations. Below are detailed descriptions of 10 of these scenarios, with foreground images taken from a participant's viewpoint data.

**Urban Arterial Road: Forward Accident - Basic**

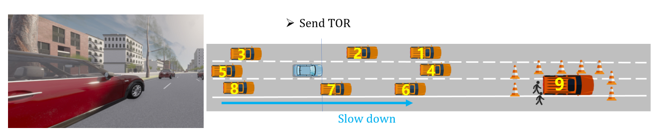


Figure S-2. Urban Arterial Road: Forward Accident - Basic

The ego vehicle is initialized in the second left lane, with a total lane length of 1500 meters, traveling at a speed of 60 km/h. In this scenario, an accident occurs ahead, causing Vehicle 4 to decelerate and stop at a rate of 7 m/s²,

ICCV #14229

while Vehicles 5, 6, 7, and 8 decelerate at 5 m/s². When the ego vehicle reaches the 1000-meter mark (36 seconds before takeover), the system issues a takeover request. The driver must assess the deceleration of the vehicles ahead and take appropriate evasive actions, such as slowing down or changing lanes. The scenario ends when the vehicle reaches the 1500-meter mark.

**Urban Arterial Road: Forward Accident - Forced Lane Change**

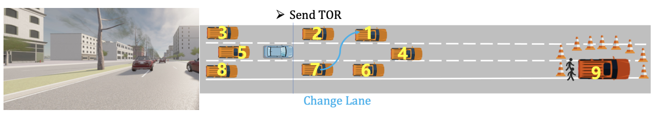


Figure S-3. Urban Arterial Road: Forward Accident - Forced Lane Change

The ego vehicle is initialized in the second left lane, traveling at 60 km/h, with a total lane length of 1500 meters. Due to an accident ahead, Vehicle 4 decelerates and stops at 7 m/s², while Vehicles 5, 6, and 8 decelerate at 5 m/s². Meanwhile, Vehicle 7 changes lanes to the leftmost lane. When the ego vehicle reaches the 1000-meter mark (36 seconds before takeover), the system issues a takeover request. The driver must avoid the accident vehicles while also monitoring the lane-changing vehicle's movement and making appropriate adjustments in speed or lane position. The scenario ends when the vehicle reaches the 1500-meter mark.

**Urban Arterial Road: Forward Accident - Accelerating Vehicles in Adjacent Lane**

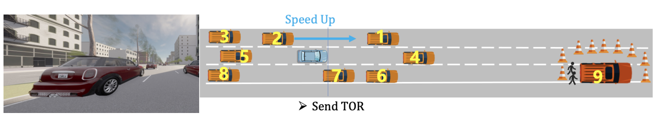


Figure S-4. Urban Arterial Road: Forward Accident - Accelerating Vehicles in Adjacent Lane

The ego vehicle is initialized in the second left lane, traveling at 60 km/h, with a total lane length of 1500 meters. Due to an accident ahead, Vehicle 4 decelerates and stops at 7 m/s², while Vehicles 5, 6, 7, and 8 decelerate at 5 m/s². At the same time, Vehicles 1, 2, and 3 accelerate to 75 km/h. When the ego vehicle reaches the 1000-meter mark (36 seconds before takeover), the system issues a takeover request. The driver must carefully assess the situation, avoiding the accident vehicles while also monitoring the rapidly accelerating vehicles behind and making appropriate decisions regarding deceleration or lane changes. The scenario ends when the vehicle reaches the 1500-meter mark.

**Urban Arterial Road: Forward Accident - Following Vehicle Changes Lanes**

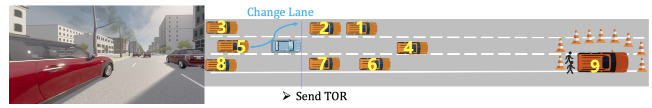


Figure S-5. Urban Arterial Road: Forward Accident - Following Vehicle Changes Lanes

The ego vehicle is initialized in the second left lane, traveling at 60 km/h, with a total lane length of 1500 meters. Due to an accident ahead, Vehicles 4, 5, 6, and 7 decelerate and stop at 5 m/s², and Vehicle 5 changes lanes to the leftmost lane. When the ego vehicle reaches the 1000-meter mark (36 seconds before takeover), the system issues a takeover request. The driver must avoid the stopped vehicles ahead while monitoring the movement of lane-changing vehicles and making appropriate driving decisions. The scenario ends when the vehicle reaches the 1500-meter mark.

**Urban Arterial Road: Forward Accident - Lead Vehicle Decelerates and Changes Lanes**

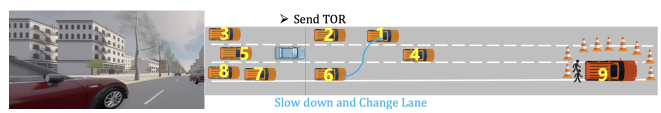


Figure S-6. Urban Arterial Road: Forward Accident - Lead Vehicle Decelerates and Changes Lanes

The ego vehicle is initialized in the second left lane, traveling at 60 km/h, with a total lane length of 1500 meters. Due to an accident ahead, Vehicles 4, 5, 7, and 8 decelerate and stop at 5 m/s², while Vehicle 6 first slows down and then changes lanes to the leftmost lane. When the ego vehicle reaches the 1000-meter mark (36 seconds before takeover), the system issues a takeover request. The driver must carefully control vehicle speed, monitor the lane-changing vehicle ahead, and make safe driving decisions. The scenario ends when the vehicle reaches the 1500-meter mark.

**Highway: Non-Motorized Vehicle Merging - Basic** The ego vehicle is initialized in the second left lane, traveling at 80 km/h, with a total lane length of 2000 meters. In this scenario, a motorcycle changes lanes multiple times and merges into the leftmost lane. When

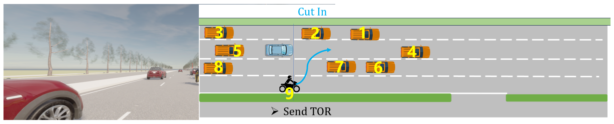


Figure S-7. Highway: Non-Motorized Vehicle Merging - Basic

the ego vehicle reaches the 1200-meter mark (40 seconds before takeover), the system issues a takeover request. The driver must assess the motorcycle's movement and take appropriate evasive actions, such as slowing down or changing lanes. The scenario ends when the vehicle reaches the 2000-meter mark.

**Highway: Non-Motorized Vehicle Merging - Overtaking Vehicle from Behind**

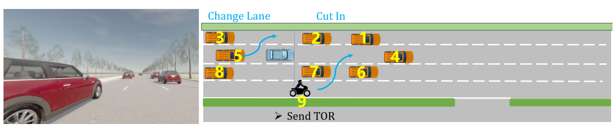


Figure S-8. Highway: Non-Motorized Vehicle Merging - Basic

The ego vehicle is initialized in the second left lane, traveling at 80 km/h, with a total lane length of 2000 meters. In this scenario, a motorcycle changes lanes multiple times and merges into the second left lane. When the ego vehicle reaches the 1200-meter mark (40 seconds before takeover), the system issues a takeover request. The driver must monitor the motorcycle's movement and make appropriate decisions regarding overtaking or maintaining a safe following distance. The scenario ends when the vehicle reaches the 2000-meter mark.

Highway: Non-Motorized Vehicle Merging - Lead Vehicle Changes Lanes

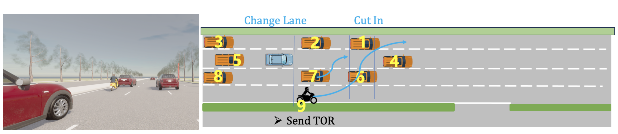


Figure S-9. Highway: Non-Motorized Vehicle Merging - Lead Vehicle Changes Lanes

The ego vehicle is initialized in the second left lane, traveling at 80 km/h, with a total lane length of 2000 meters. A motorcycle is traveling ahead of the ego vehicle and changes lanes multiple times before merging into the leftmost lane. When the ego vehicle reaches the 1200-meter mark (40 seconds before takeover), the system issues a takeover request. The driver must assess the motorcycle's movement and adopt an appropriate driving strategy, such as slowing down or changing lanes to avoid a collision. The scenario ends when the vehicle reaches the 2000-meter mark.

**Highway: Non-Motorized Vehicle Merging - Sudden Cut-In from Adjacent Vehicle**

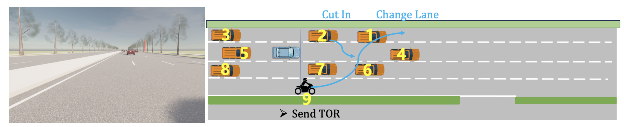


Figure S-10. Highway: Non-Motorized Vehicle Merging - Sudden Cut-In from Adjacent Vehicle

The ego vehicle reaches the 1200-meter mark (40 seconds before takeover), the system issues a takeover request. The driver must quickly recognize the motorcycle's cut-in maneuver and take appropriate evasive actions, such as slowing down or changing lanes. The scenario ends when the vehicle reaches the 2000-meter mark.

**Highway: Non-Motorized Vehicle Merging - Sudden Stop of Lead Vehicle** The ego vehicle is initialized

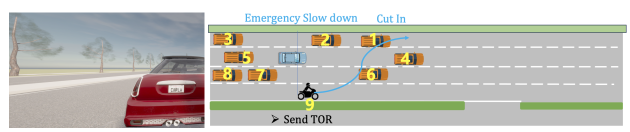


Figure S-11. Highway: Non-Motorized Vehicle Merging - Sudden Stop of Lead Vehicle

in the second left lane, traveling at 80 km/h, with a total lane length of 2000 meters. In this scenario, a motorcycle changes lanes multiple times and merges into the leftmost lane, while at the same time, Vehicle 4, traveling ahead of the ego vehicle, suddenly decelerates to 5 km/h. When the ego vehicle reaches the 1200-meter mark (40 seconds before takeover), the system issues a takeover request. The driver must avoid the motorcycle while quickly assessing the situation with the suddenly stopped vehicle ahead and take emergency measures, such as hard braking or lane changing, to ensure safety. The scenario ends when the vehicle reaches the 2000-meter mark.

ICCV #14229

| Index | Feature Name | Input to Model | Used in Observation Model | Used for Sketches Synthesis | Activated in Final Rules |
|---|---|---|---|---|---|
| 1 | steer | √ | | √ | |
| 2 | acc | √ | | √ | |
| 3 | brake | √ | | √ | √ |
| 4 | fixation obj distance | √ | | √ | √ |
| 5 | fixation obj speed | √ | | √ | √ |
| 6 | fixation obj yaw | √ | | √ | √ |
| 7 | left obj distance | √ | | √ | √ |
| 8 | left obj speed | √ | | √ | √ |
| 9 | left obj yaw | √ | | √ | |
| 10 | right obj distance | √ | | √ | √ |
| 11 | right obj speed | √ | | √ | √ |
| 12 | right obj yaw | √ | | √ | |
| 13 | ego vehicle speed | √ | | √ | √ |
| 14 | ego vehicle yaw | √ | | √ | √ |
| 15 | gaze point x (screen) | √ | √ | | |
| 16 | gaze point y (screen) | √ | √ | | |
| 17 | fixation obj x (screen) | √ | √ | | |
| 18 | fixation obj y (screen) | √ | √ | | |
| 19 | left obj x (screen) | √ | √ | | |
| 20 | left obj y (screen) | √ | √ | | |
| 21 | right obj x (screen) | √ | √ | | |
| 22 | right obj y (screen) | √ | √ | | |

Table S-2. Definitions and Usage of Features in the PILOT Model

## S-2. Supplementary Information on the PILOT Model

### S-2.1. Features used for sketch synthesis

We incorporated 22 features into the model, with 14 of them used to generate strategy sketches, constructing 93 fundamental sketch structures. These sketches were further combined using logical connectors (AND, OR, NOT) to create a diverse rule model space. The training results indicate that among the 14 features utilized for sketch synthesis, 10 were actively involved in the final rule generation process. The table S-2 below presents the features and their usage details.

### S-2.2. PILOT mutations

Our incremental heuristic search method allows for the following six types of mutations in the policy sketches:
(1) Add a new threshold predicate with a random feature.
(2)Simplify the policy by removing an existing predicate.
(3) Swap conjunctions with disjunctions, and vice-versa.
(4) Augment numerical features by applying random functions with random parameters.
(5) Simplify features by removing function applications.
(6) Reset to simpler policies For instance

$$
\begin{aligned}
&\text{speed} > 20 \xrightarrow{(1)} \text{speed} > 20 \wedge \text{yaw} < 15 \xrightarrow{(3)} \\
&\text{speed} > 20 \vee \text{yaw} < 15 \xrightarrow{(2)} \text{yaw} < 15 \xrightarrow{(4)} \\
&\text{yaw} + \text{left\_yaw} < 15 \xrightarrow{(5)} \text{left\_yaw} < 15 \\
&\xrightarrow{(6)} \text{speed} > 20
\end{aligned}
$$

## References

[1] Pushpa Choudhary, Akshay Gupta, and Nagendra R Velaga. Perceived risk vs actual driving performance during distracted driving: A comparative analysis of phone use and other secondary distractions. *Transportation research part F: traffic psychology and behaviour*, 86:296–315, 2022. 2

[2] Jianwu Fang, Dingxin Yan, Jiahuan Qiao, Jianru Xue, He Wang, and Sen Li. Dada-2000: Can driving accident be predicted by driver attention$f$ analyzed by a benchmark. In *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, pages 4303–4309. IEEE, 2019. 1

[3] R. Kohli. Assessing interaction effects in latin square-type

designs. *International Journal of Research in Marketing*, 5: 25–37, 1988. 2

[4] Lingqiao Qin, Zhixia Li, Zhijun Chen, M.S. Andi Bill, and D. Noyce. Understanding driver distractions in fatal crashes: An exploratory empirical analysis. *Journal of safety research*, 69: 23–31, 2019. 2

[5] Dongjian Song, Bing Zhu, Jian Zhao, and Jiayi Han. Human–machine shared lateral control strategy for intelligent vehicles based on human driver risk perception reliability. *Automotive Innovation*, 7(1):102–120, 2024. 2

[6] D. Strayer. Is the technology in your car driving you to distraction? *Policy Insights from the Behavioral and Brain Sciences*, 2:157 – 165, 2015. 2

[7] Ye Xia, Danqing Zhang, Jinkyu Kim, Ken Nakayama, Karl Zipser, and David Whitney. Predicting driver attention in critical situations. In *Computer vision–ACCV 2018: 14th Asian conference on computer vision, perth, Australia, December 2–6, 2018, revised selected papers, part v 14*, pages 658–674. Springer, 2019. 1