Human-likeness Assessment of Driving Performance for Intelligent Vehicles

Chen Wang, Ph.D.,

Professor of Intelligent Transportation Systems

Southeast University

15005172397

chen_david_wang@seu.edu.cn.

About Me



Chen Wang

Chair Professor
Director of International Affairs
Deputy Director of Intelligent Mobility Research Center
Southeast University

Research interests: Intelligent Mobility, Autonomous Vehicles, Multi-modal Perception, including 2D/3D Vehicle Detection, Road Anomaly/Incident Detection, Multimodal Large Models, etc.

Educational Background: Bachelor and Master Degree (Southeast University, China); PH.D Degree (University of Kentucky, USA)

Work Experience: Tongji University(Assistant Professor) and Southeast University(Professor)

Research Papers: 100+ papers published, including IEEE TII/TITS/IOTJ/TR PART C/AAP etc./World's Top 2% Scientists by Stanford and Elsevier

Research Projects: lead 30+ research projects with total funding over 9 million USD

Academic Appointments:

Associate Editor of *IEEE Transactions on Intelligent Transportation Systems*/ *IEEE Open Journal of Intelligent Transportation Systems*

Academic Promotion:

WeChat public account " Deep Traffic" with over 10k followers

DEEP TRAFFIC

Outline

1.Background and Motivation

2.Methods and Experiments

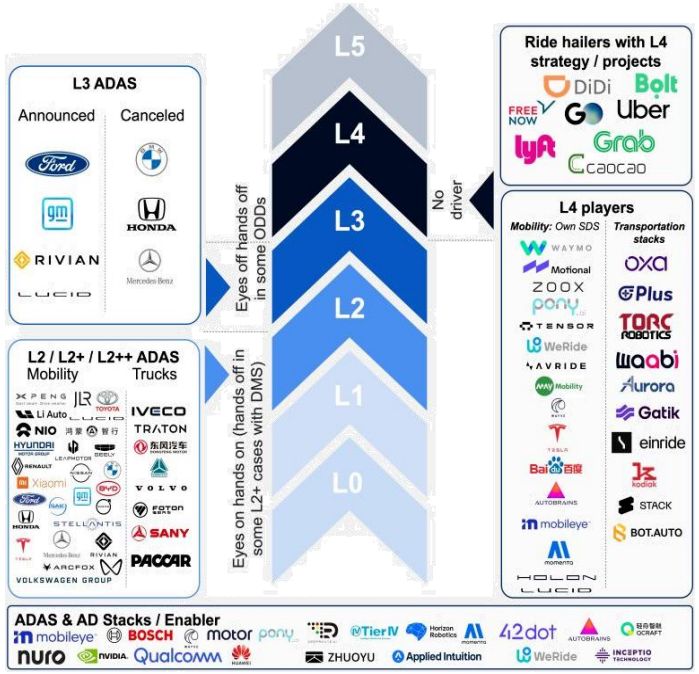
3.Results and Findings

4.Future Prospects

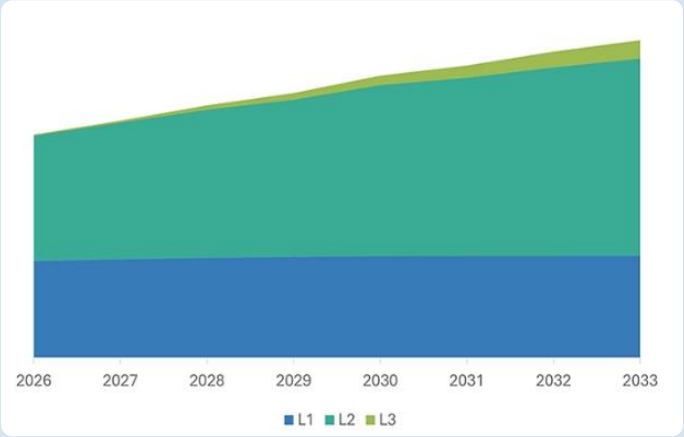
Background and Motivation

Driven by rapid advances in high-level ADAS and automated driving, intelligent vehicles are moving from pilot deployment toward large-scale commercialization across major automotive markets worldwide.

Industry Development



ADAS MARKET, BY LEVEL OF AUTONOMY, 2026-2033 (THOUSAND UNITS)

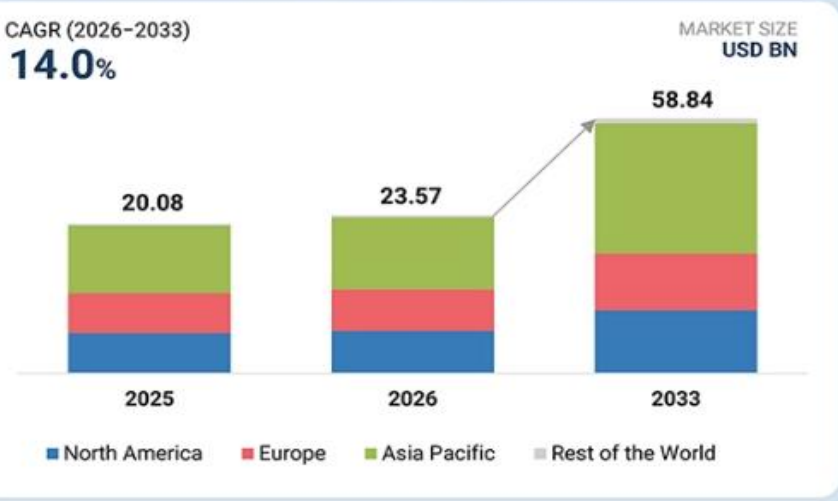


Global OEMs accelerates the deployment of high-level automated driving technologies (L2-L4) across multiple mobility sectors.

L3 segment to grow at fastest rate in ADAS market during forecast period.

Market Size and Growth Trends

ADAS (COMPONENT): MARKET SIZE & SHARE



Market Size in 2025	USD 20.08 BN
Market Forecast in 2033	USD 58.84 BN
CAGR	14.0%

Source:. arketsandMarkets, Advanced Driver Assistance Systems (ADAS) Market, 2026; MarkLines, PwC Research and Analysis, 2026; MarketsandMarkets, Advanced Driver Assistance Systems (ADAS) Market, 2026..

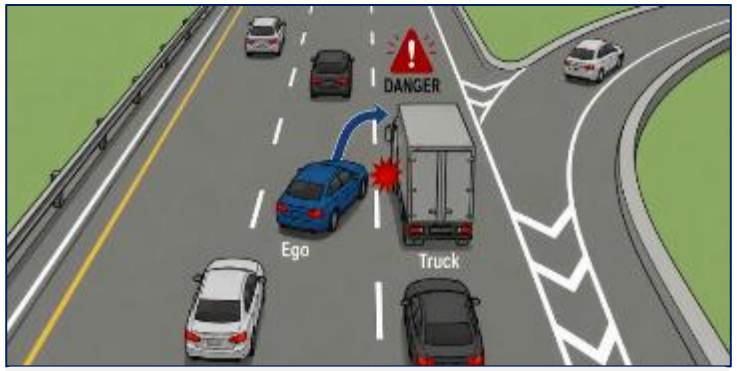
Background and Motivation

GAP: Purely machine-centric driving behaviors lack similarity to competent and careful human drivers, which could lead to unsafe, uncomfortable and inefficient performance in complex traffic environments.

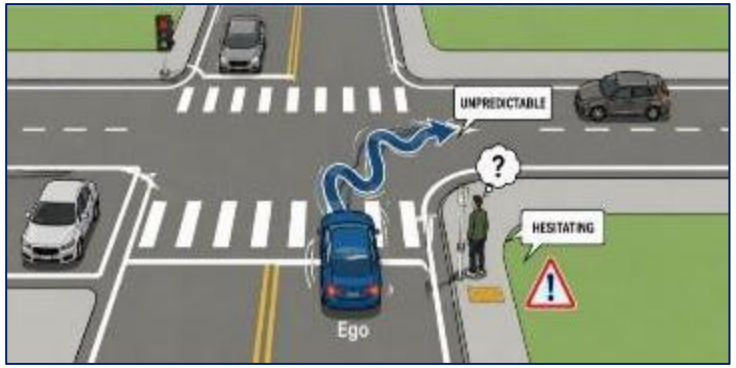
Typical Machine-centric Driving



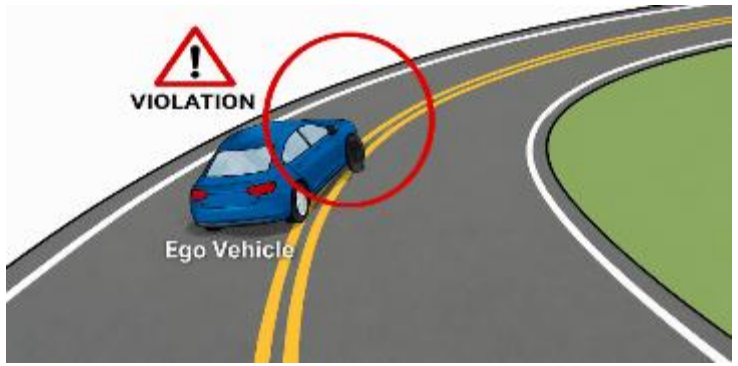
A. Over-conservative decision-making



B. Risky overtaking

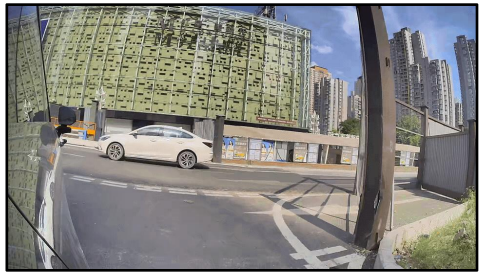


C. Incompatible unstable control



D. Violating traffic rules

Negative Impacts



Increased Interaction Risks



Reduced Traffic Efficiency

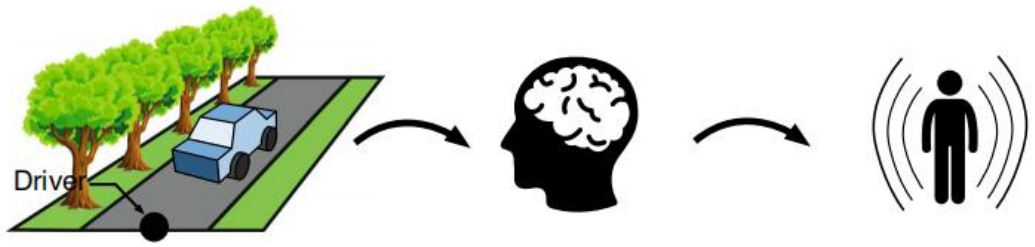


Reduced User Trust and Acceptance

Background and Motivation

TREND: Human-like driving has become an important trend in the development of intelligent vehicles

Risk perception and decision-making process



What is human-like driving?



Human-like driving behavior refers to driving behavior that emerges from models reproducing human risk perception and decision-making characteristics.

Growing International Attention

Increased Attention

Human Benchmark

Human driving data are used as benchmarks for ADS behavioral competence evaluation.

Behavior Requirement

ADS should behave comparably to a competent and careful human driver.

Assessment Needs

Human-like driving assessment has become a national strategic needs.

Source: *Kolekar S, De Winter J, Abbink D. Human-like driving behaviour emerges from a risk-based driver model[J]. Nature communications, 2020, 11(1): 4850.*

Background and Motivation

INITIATIVE: China's Ministry of Industry and Information Technology stress the need to evaluate Intelligent Vehicles against human-like driving benchmarks.


中华人民共和国工业和信息化部
Ministry of Industry and Information Technology of the People's Republic of China

首页 组织机构 新闻发布 政务公开 政务服务 互动交流 工信数据

首页 > 工业和信息化部 > 机关司局 > 装备工业一司 > 汽车工业

公开征求《智能网联汽车 自动驾驶系统安全要求》等五项强制性国家标准的意见

发布时间：2026-02-12 18:45 来源：装备工业一司

按照《中华人民共和国标准化法》和《强制性国家标准管理办法》，工业和信息化部装备工业一司组织全国汽车标准化技术委员会开展了《智能网联汽车 自动驾驶系统安全要求》等五项强制性国家标准的制修订，已形成征求意见稿，现公开征求社会各界意见。

征求意见截止日期为2026年4月13日，如有意见和建议，请以书面（个人需署名，单位需加盖公章，并留联系方式）或电子邮件形式进行反馈。

附件：

1. 智能网联汽车 自动驾驶系统安全要求（征求意见稿）.pdf
2. 智能网联汽车 自动驾驶系统安全要求（征求意见稿）编制说明.pdf
3. 汽车操纵件、指示器及信号装置的标志（征求意见稿）.pdf
4. 汽车操纵件、指示器及信号装置的标志（征求意见稿）编制说明.pdf
5. 轻型汽车制动辅助系统（BAS）性能要求及试验方法（征求意见稿）.pdf
6. 轻型汽车制动辅助系统（BAS）性能要求及试验方法（征求意见稿）编制说明.pdf
7. 摩托车和轻便摩托车后视镜的性能和安装要求（征求意见稿）.pdf
8. 摩托车和轻便摩托车后视镜的性能和安装要求（征求意见稿）编制说明.pdf
9. 低地板及低入口城市客车结构要求（征求意见稿）.pdf
10. 低地板及低入口城市客车结构要求 （征求意见稿）编制说明.pdf
11. 强制性国家标准反馈意见表.doc

联系方式：

"The safety performance of ADAS shall be at least equivalent to that of a ***competent and attentive human driver.*** *"*


ICS 43.020
CCS T 40

GB

ICS 42.020
CCS T40

GB

中华人民共和国国家标准

GB XXXX—XXXX
代替 GB/T 44721—2024

智能网联汽车 自动驾驶系统安全要求

Intelligent and connected vehicle — Safety requirements for automated driving system

（征求意见稿）

（本草案完成时间：2026 年 2 月 5 日）

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。

XXXX－XX－XX 发布 XXXX－XX－XX 实施

国家市场监督管理总局
国家标准化管理委员会 发布

Existing proposals provide the general concept without concrete, usable rules for benchmark, scenario design, performance metrics, and assessment methods.

Outline

1.Background and Motivation

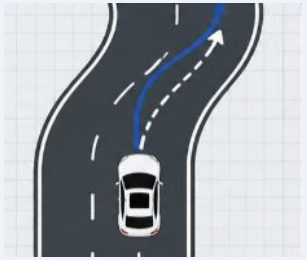
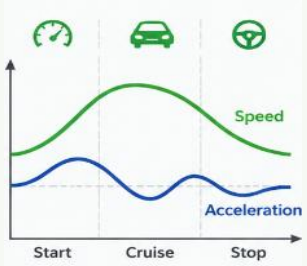
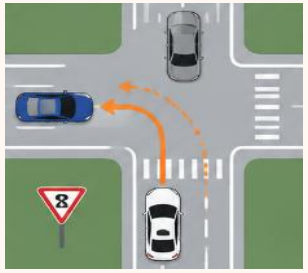
2.Methods and Experiments

3.Results and Findings

4.Future Prospects

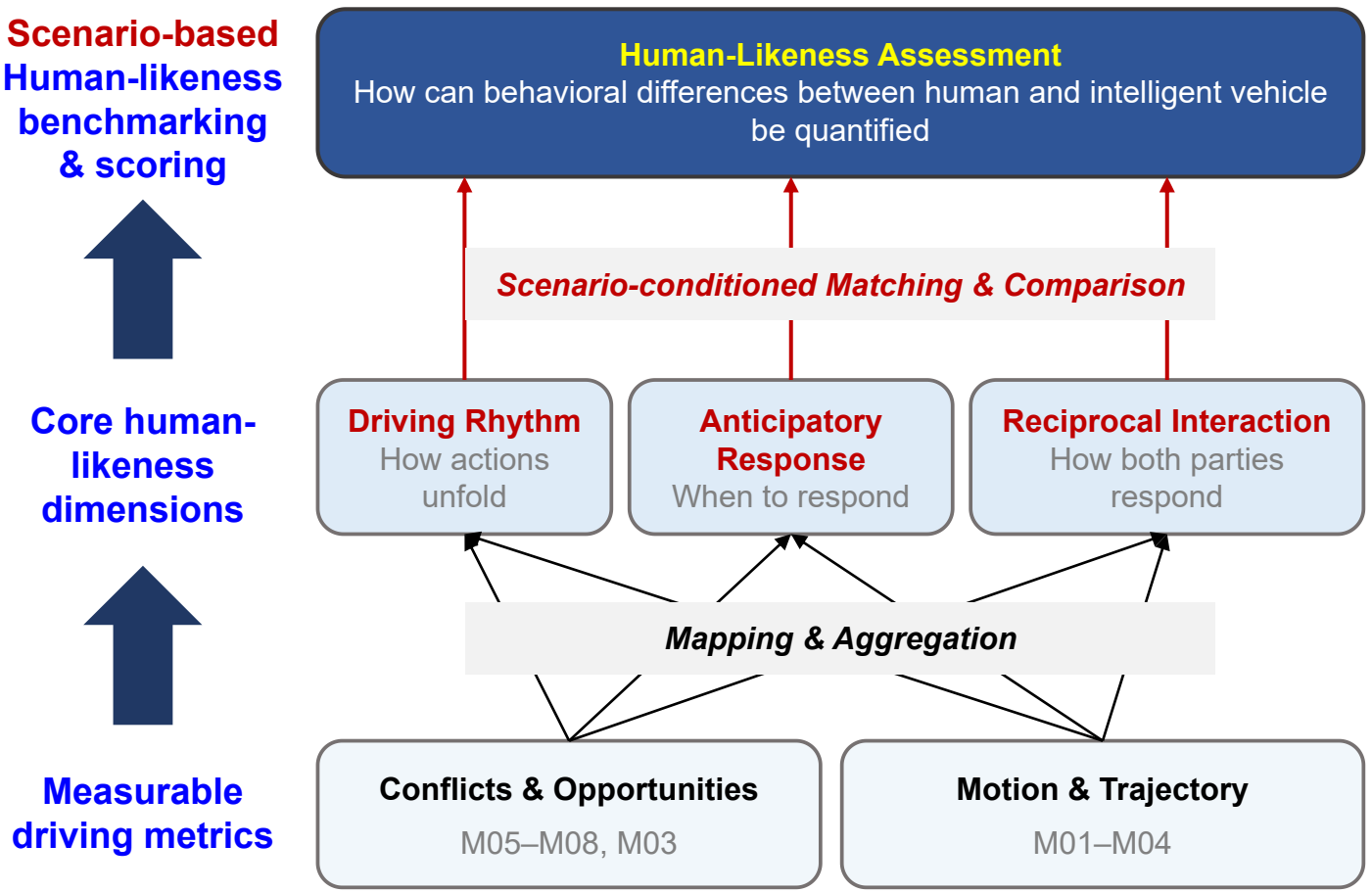
Methods and Experiments

KEY CONCEPTS OF HUMAN-LIKENESS

Human-likeness Dimension		Criterion
	**Driving Rhythm (DR)**	How smoothly the ego vehicle's **longitudinal and lateral control** actions start, evolve, and settle, evaluated independently of when they are triggered and of how other vehicles behave. $\{ACC_{ego}, \ \mathrm{Yaw}_{ego}, Jerk_{ego}, Lateral_offset_{ego}\}$
	**Anticipatory Response (AR)**	How early the ego vehicle initiates a maneuver relative to the evolving traffic event, captured by **timing measures**. $\{TTC_{ego}, \ \min_Gap, \ ACC_{ego}, Jerk_{ego}, \ ACC_{non_ego}, R_{delay}\}$
	**Reciprocal Interaction (RI)**	How the ego vehicle and its interaction partner mutually adjust their motions, captured by **relative measures** such as who reacts first and how closely their adjustments mirror each other. $\{Risk_{ego}, \ Risk_{non_ego}, T_{loss}, \ ACC_{ego}, ACC_{non_ego}\}$

Methods and Experiments

HUMAN-LIKENESS ASSESSMENT : A hierarchical assessment architecture links measurable driving metrics to core human-likeness dimensions and, ultimately, to human-likeness benchmarking and rating.



Measurable driving metrics definitions

Metric ID	Definition	Value
M01	Ego resultant acceleration RMS (ACC_{ego})	Mean (↓)
M02	Ego yaw-rate RMS (Yaw_{ego})	Mean (↓)
M03	Longitudinal jerk rate ($Jerk_{ego}$)	Max (↓)
M04	Ego Lateral Off set ($Lateral_offset_{ego}$)	Max (↓)
M05	Two-dimensional time to collision (TTC_{ego})	min (↑)
M06	Response delay (R_{delay})	— (↓)
M07	Interacting-vehicle TTC-based aggregated conflict risk index ($Risk_{ego}$)	Max (↓)
M08	Travel Time Loss (T_{loss})	— (↓)

Methods and Experiments

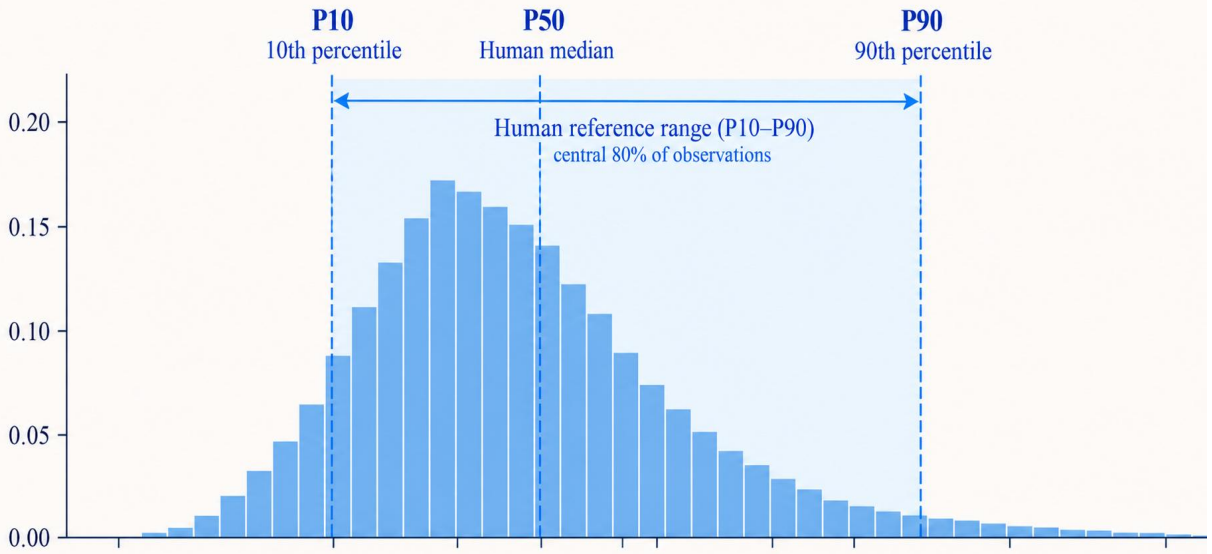
HUMAN-LIKENESS ASSESSMENT : Derive scenario-based human benchmark from naturalistic driving trajectory data, then assess intelligent vehicle performance by comparing it with human benchmark.

BUILD THE HUMAN BENCHMARK

Empirical human driving reference

Example of one dimension:

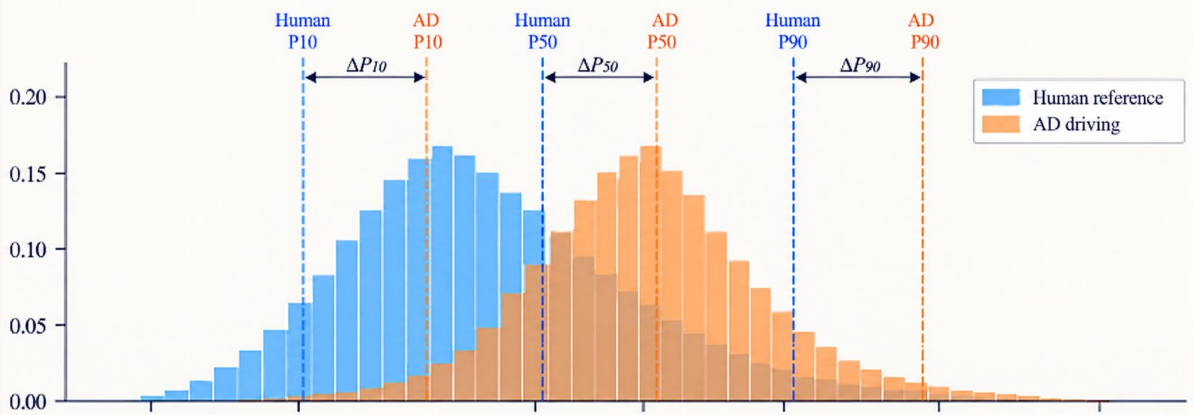
$$\boldsymbol{Driving\ Rhythm} = \boldsymbol{f}_{(ACC_{ego},\ Yaw_{ego},\ Jerk_{ego},\ Lateral_offset_{ego})}$$



Human driving parameters: $B^{H}_{c,b} = (P_{10}, P_{50}, P_{90})$

QUANTIFY HUMAN-LIKENESS

Behavioral Deviation Quantification



Driving Rhythm Similarity Score (S_{DR})

$$\Delta P_{q_i} = \frac{|P^{AD}_{q_i} - P^{H}_{q_i}|}{\max(P^{H}_{90} - P^{H}_{10}, \delta)},\ i = 1, \ldots, N$$

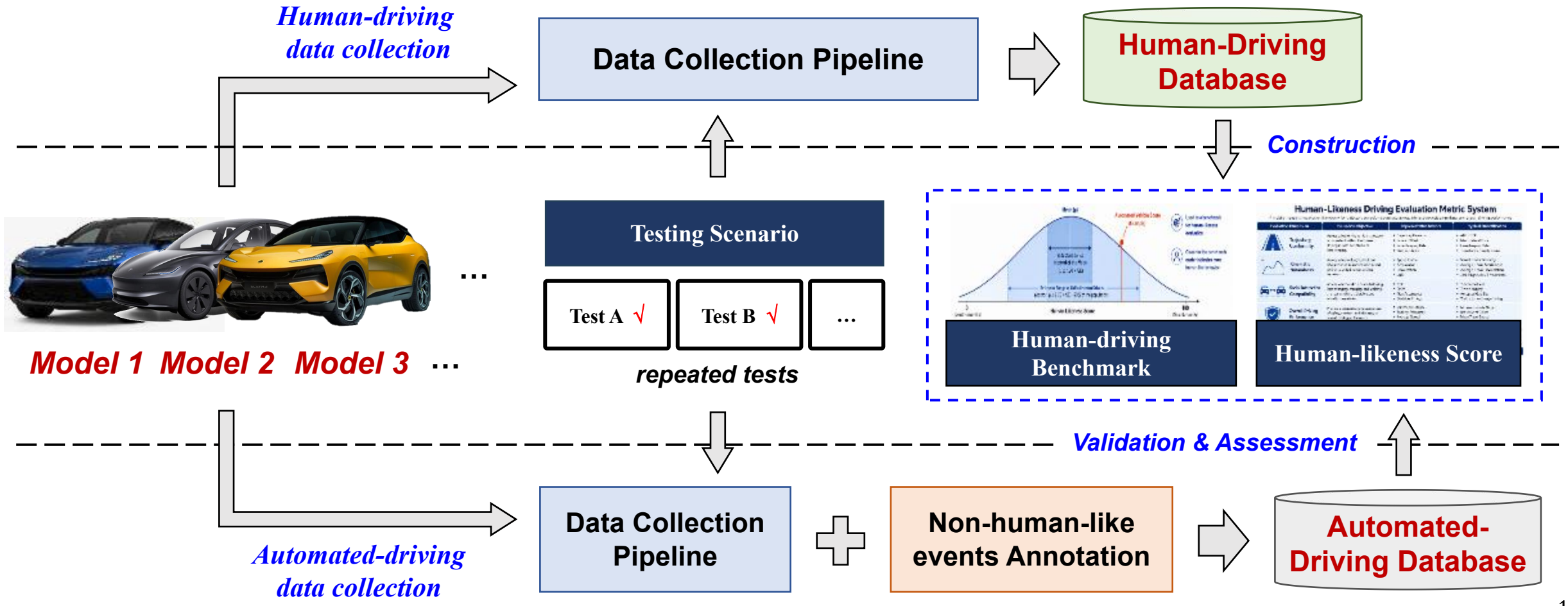
$$S_{DR} = \exp\left(-\frac{1}{N}\sum_{i=1}^{N}\Delta P_{q_i}\right),\ 0 < S_{DR} \le 1$$

Score for full dimensions

$$S_{scene} = \sqrt[3]{S_{DR} \times S_{AR} \times S_{RI}}$$

Methods and Experiments

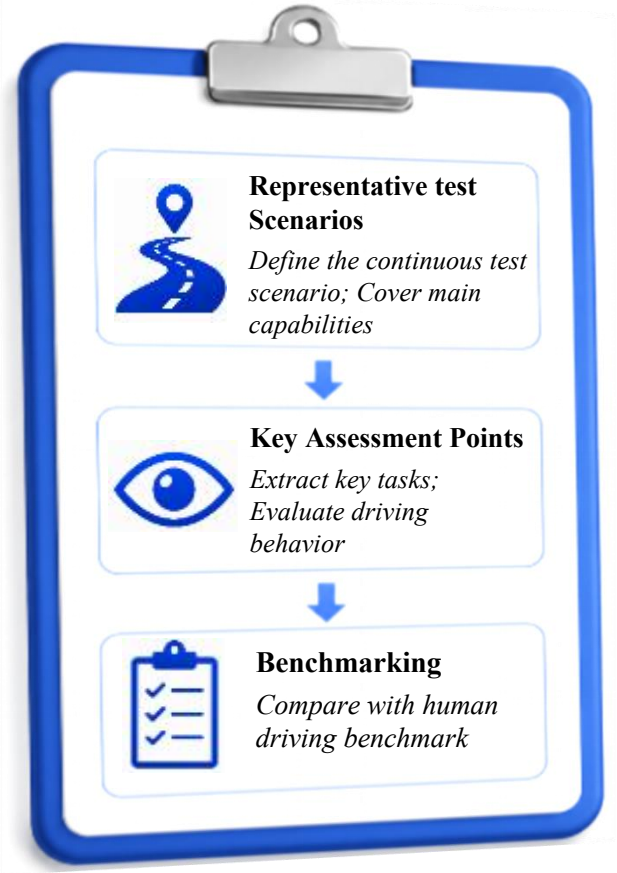
EXPERIMENTAL DESIGN: A unified scenario-based testing framework integrates human- and automated-driving data collection, human-driving benchmark construction, non-human-like event annotation, and human-likeness assessment and validation.



Methods and Experiments

COMPOSITE ROAD TEST SCENARIO DESIGN:

Design scenario-based tests which embed key driving tasks in a complex traffic environment and targets specific human-likeness dimensions.



Test A

Scenario: Highway continuous S-curve;

Test Tasks: *Curve lane-keeping & Ramp curvature transitions*

Assessment Focus: *Driving Rhythm*

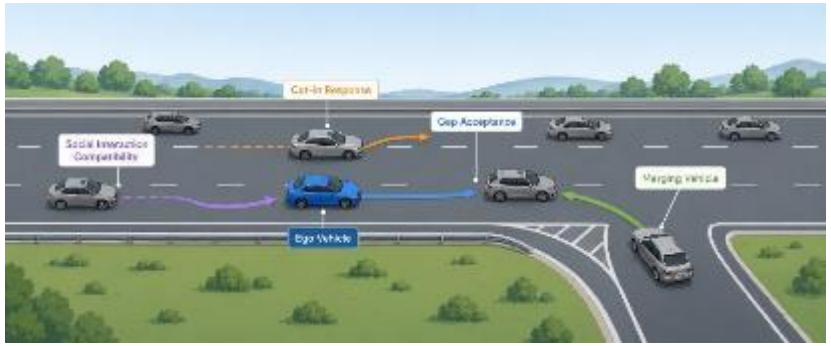


Test B

Scenario: Weaving area; stop-and-go traffic

Test Tasks: *Ramp merging & Cut-in / Cut-off & Weaving interaction*

Assessment Focus: *Driving Rhythm, Anticipatory Response, Reciprocal Interaction*

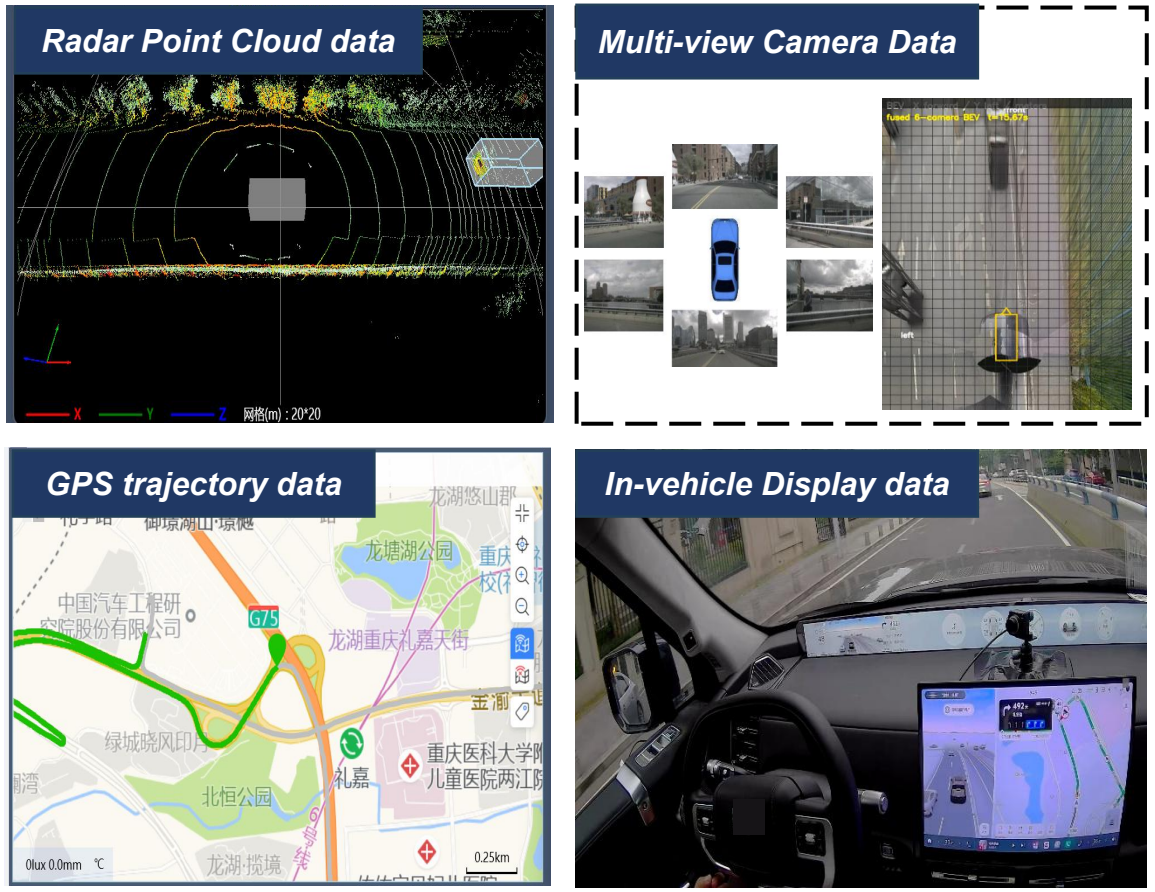


Methods and Experiments

DATA COLLECTION: Hardware and Software Platform (i-DAF from CAERI)



Equipment Deployment Example



Multi-modal / Multi-view Data Collection

Methods and Experiments

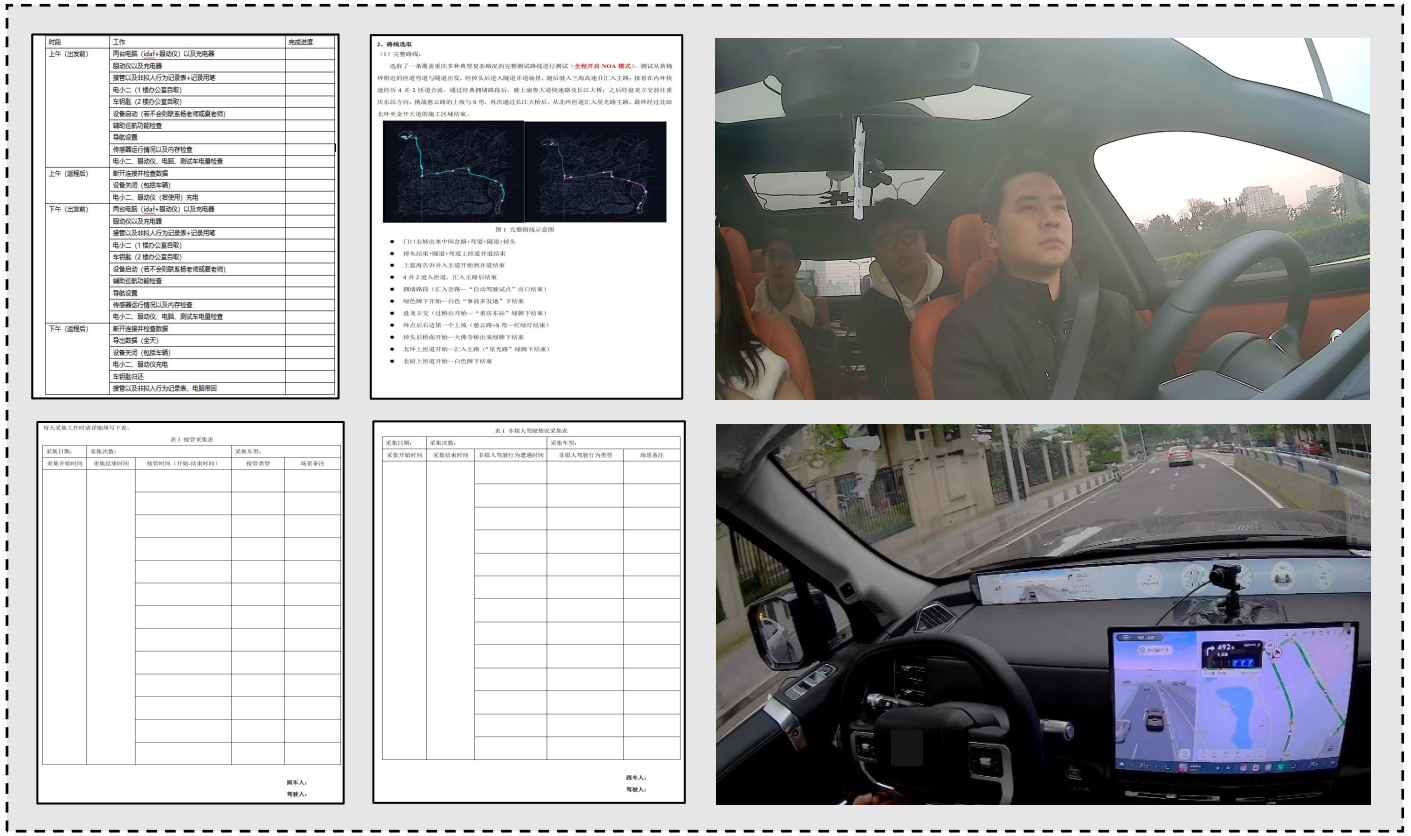
DATA COLLECTION: Collect and annotate driving behaviors of intelligent vehicles deviating from those of competent and careful human drivers in scenario-based road tests

How:

- **One-month** data collection
- **3** vehicle models, **8** participants
- Manually document and **cross-check** intelligent vehicle responses and driver takeover events
- Timestamp and annotate **observable non-human-like driving events**

Output:

- More than **10 TB multi-source** data
- Representative non-human-like driving events and **structured taxonomy**
- **Annotated database** of deviation and driver takeover events

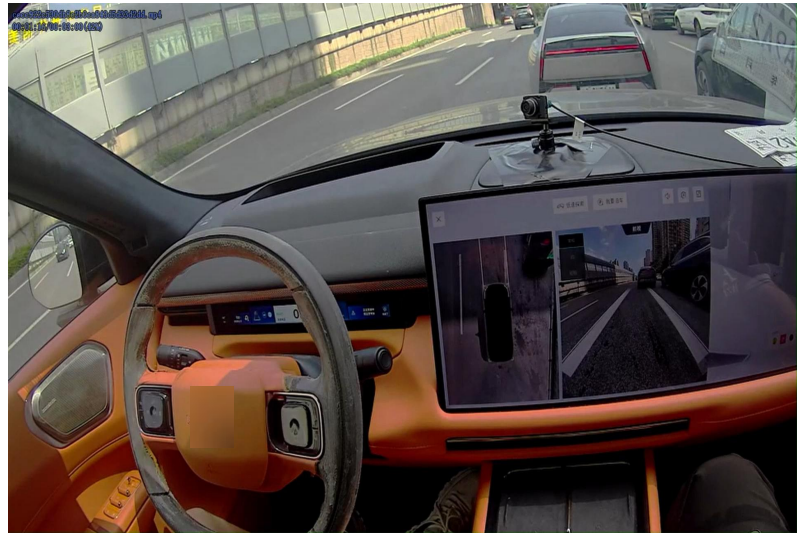


Standardized documentation forms for recording non-human-like driving behaviors and takeover events

Methods and Experiments

DATA COLLECTION: Document Non-Human-Like Driving events for analysis and validation

Key Question	Characterization Dimension	Typical Labels
Where/ When	Identify the road environment and driving task in which the non-human-like behavior/takeover occurred.	① S-curve, ② Roundabout, ③ On-ramp/Off-ramp, ④ Urban multilane road, etc.
What	Describe the observable behavior that deviated from skilled human driving.	① Unstable following or lane keeping ② Inappropriate lane change/overtaking ③ Avoidance (obstacles, other vehicles, pedestrians etc.) anomalies ④ Start / Stop anomalies, etc.
Why	Attribute the observed behavior to the primary automated-driving capability involved.	① Perception deficiencies (scenario understanding) ② Planning deficiencies (intent prediction) ③ Control deficiency (Vehicle response)



Case A:
Risky Lane-Change Event

- **Where/when:** Urban multilane road
- **What:** Abrupt lane change
- **Why:** Planning deficiency
- **Intervention:** Driver takeover

Methods and Experiments

Scenario-based Human-Likeness Assessment Standard and Certification



Model 1 Model 2 Model 3 ...

Select Target Test Scenarios

Apply the Scenario-Matched Human-Driving Benchmark

Human-Likeness Rating



A | B | C | D | E

Illustrative Rating Results

Vehicle	Test A	Test B	Rating	Certification
Vehicle 1	A	A	A	Certified
Vehicle 2	B	B	B	Certified
Vehicle 3	C	C	C	Certified
Vehicle 4	D	D	D	Certified
.....	E	E	E	Not Certified

A Excellent | B Good | C Acceptable | D Marginal | E Poor

Outline

1.Background and Motivation

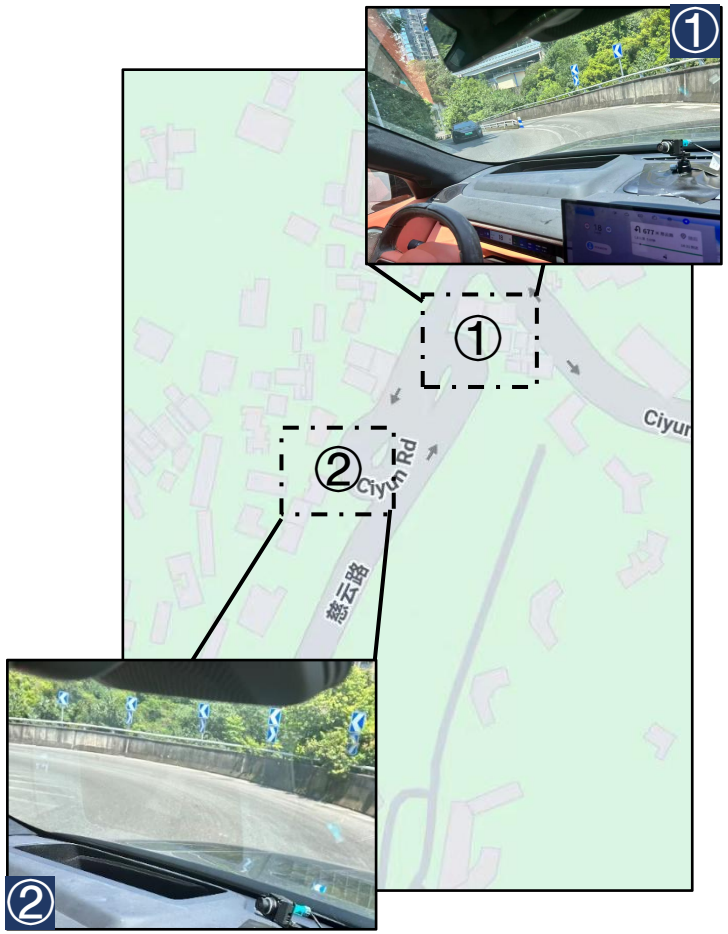
2.Methods and Experiments

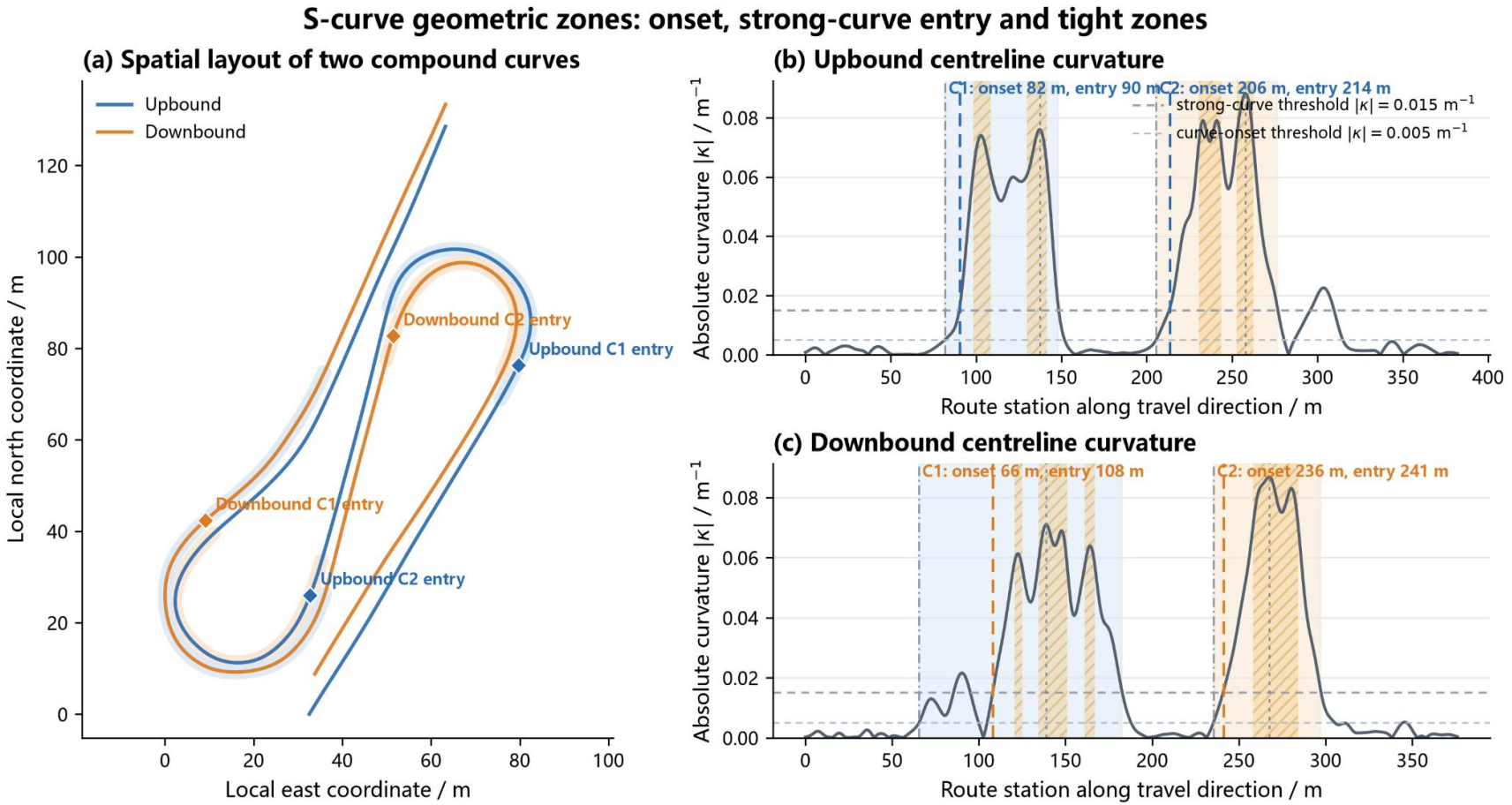
3.Results and Findings

4.Future Prospects

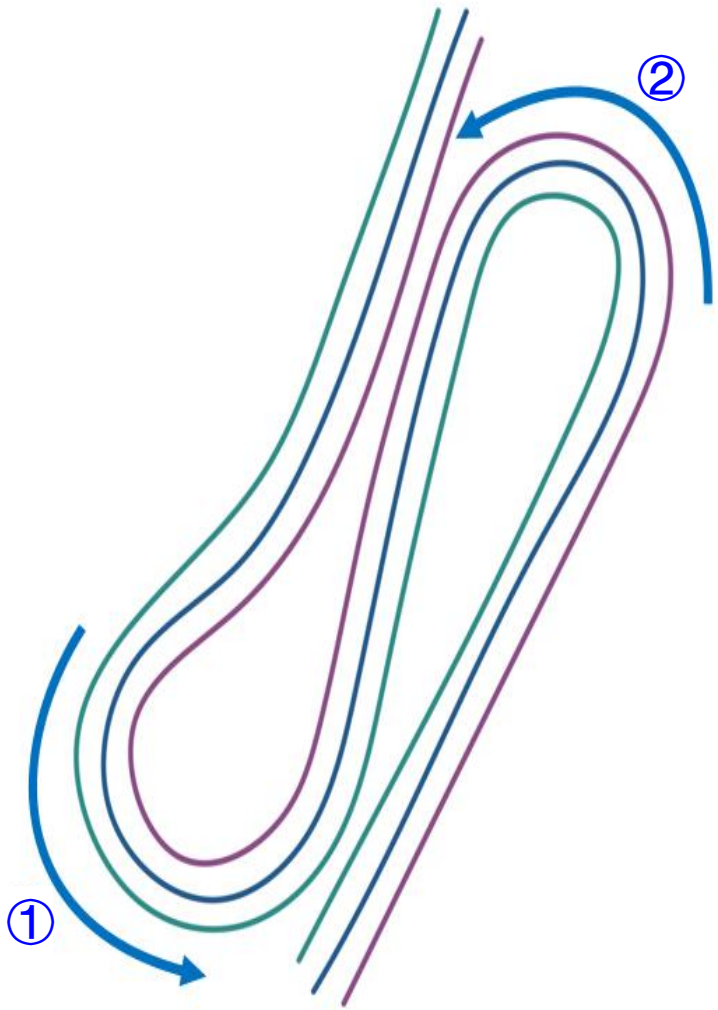
Results and Findings

TEST A: Continuous curvature variations and constrained lateral space make it challenging for vehicles to maintain a smooth driving rhythm during motion planning and lane keeping.





Results and Findings



Total Mileage

90.2 km

3 vehicle models, **14** repeated rounds per model

Non-Human-Like Driving Events

220

2.44 events per kilometer

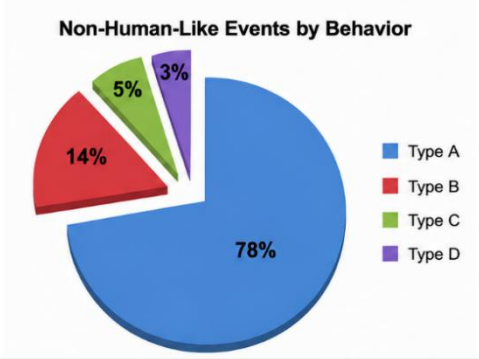
Trigger

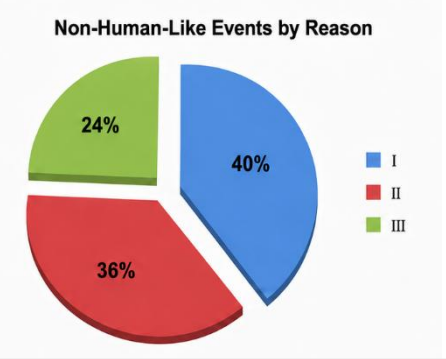
Takeovers

73

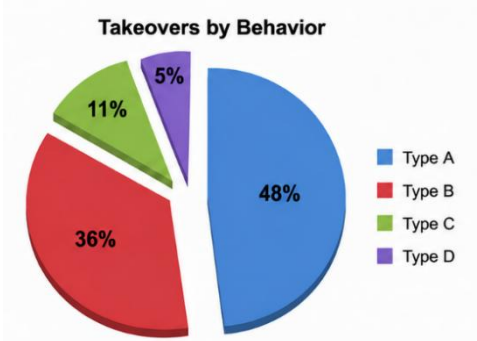
0.81 events per kilometer

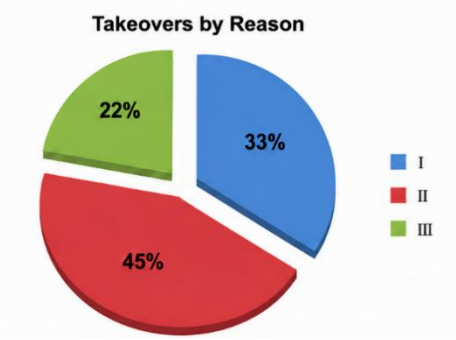
Non-Human-Like Driving Events





Takeovers





Abnormal Behavior Types

A. Unstable following or lane keeping

B. Inappropriate lane change/overtaking

C. Avoidance (obstacles, other vehicles, pedestrians etc.) anomalies

D. Start / Stop anomalies, etc.

Behavior Reason Annotation

I. Inaccurate Perception

II. Improper Planning

III. Incorrect Response

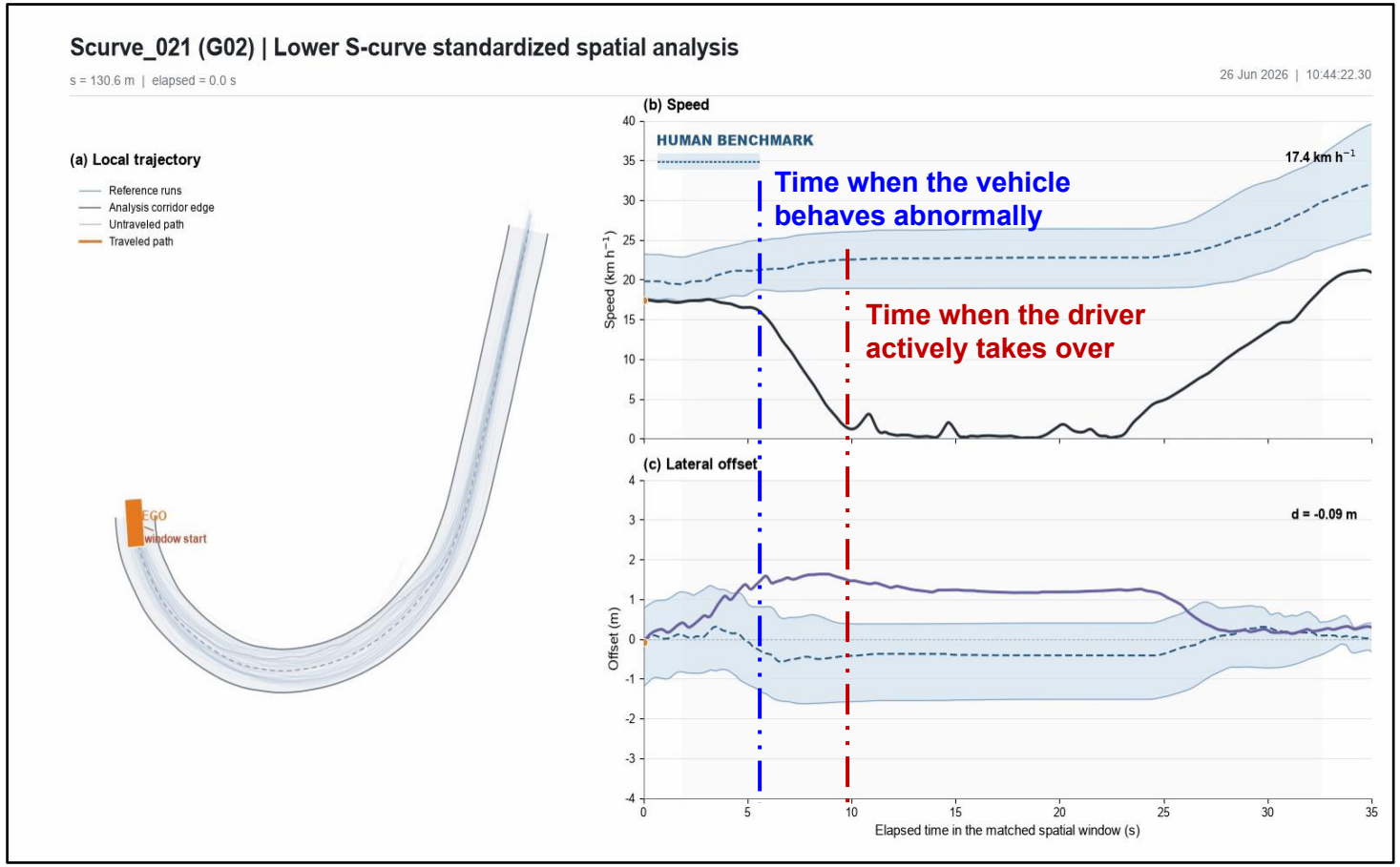
Results and Findings

Case Study (Point ①) : Active Takeover

The AD exhibited significant path deviation, entering the oncoming lane and ultimately leading to active takeover.



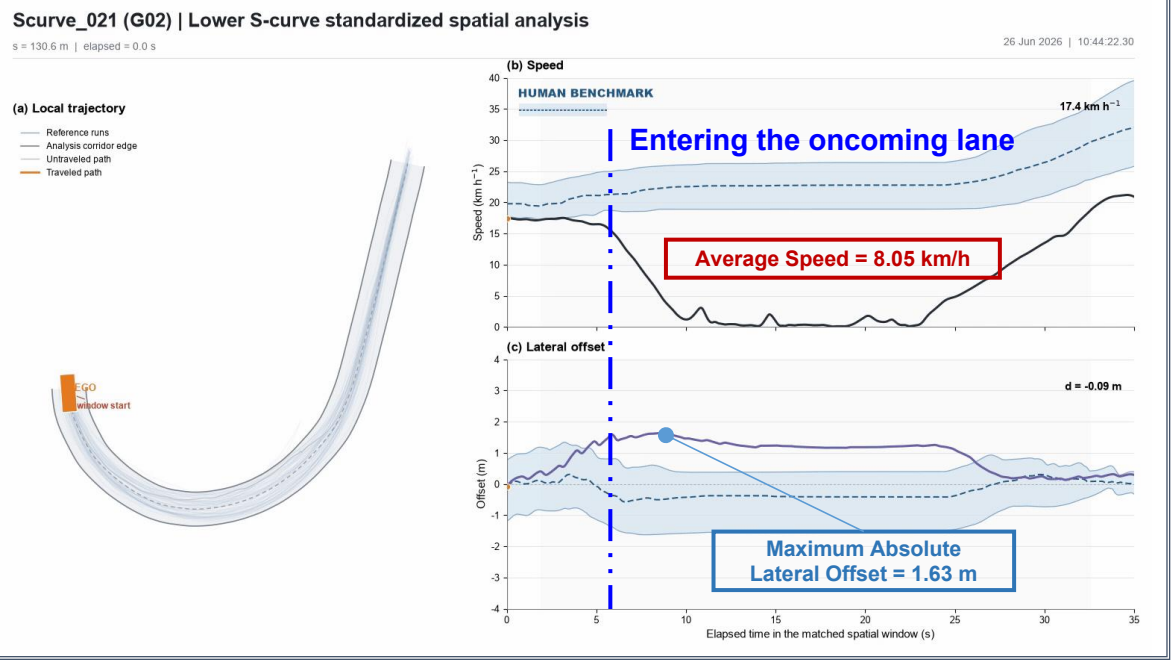
- **Where/when:** S-curve
- **What:** Unstable lane keeping
  (entering the oncoming lane)
- **Why:** Improper planning and control
- **Intervention:** Driver active takeover



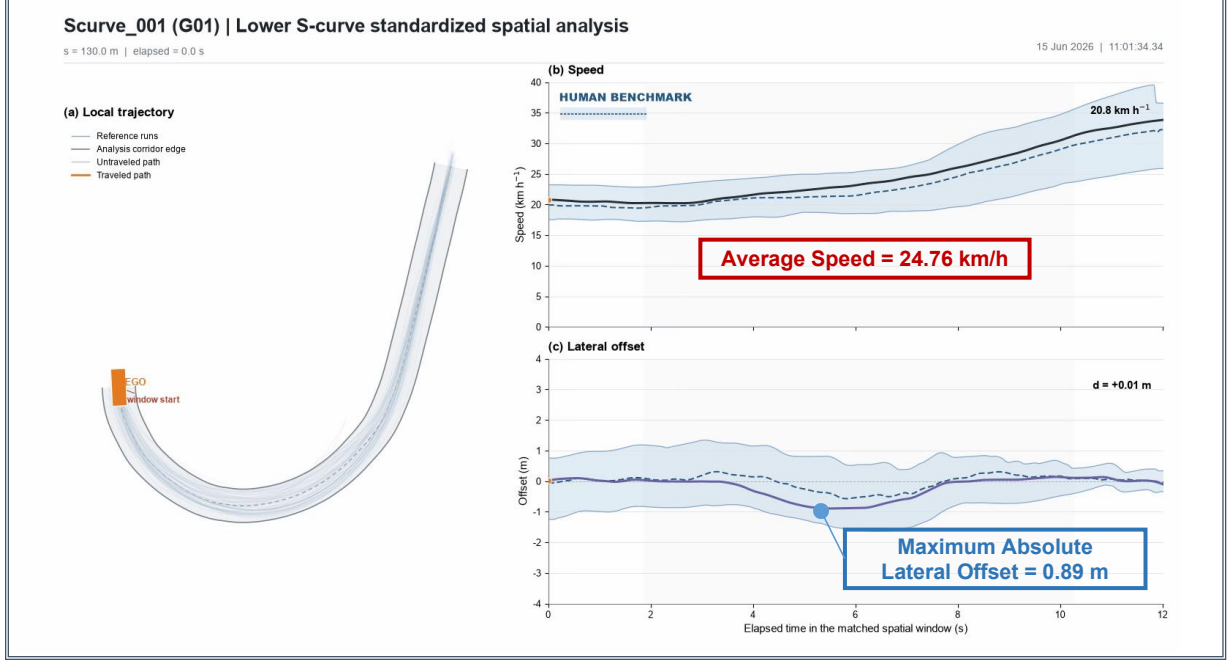
Results and Findings

Case Study (Point ①) : Active Takeover

Non-Human-Like Driving



Human-Like Driving



	Average Speed (km/h)	Maximum Absolute Lateral Offset (m)
Non-Human-Like Driving	8.05	1.63
Human Benchmark	20.20 (17.22, 39.53)	1.34 (0, 1.34)

Non-human-like driving in S-curve: unexpected large lateral offset

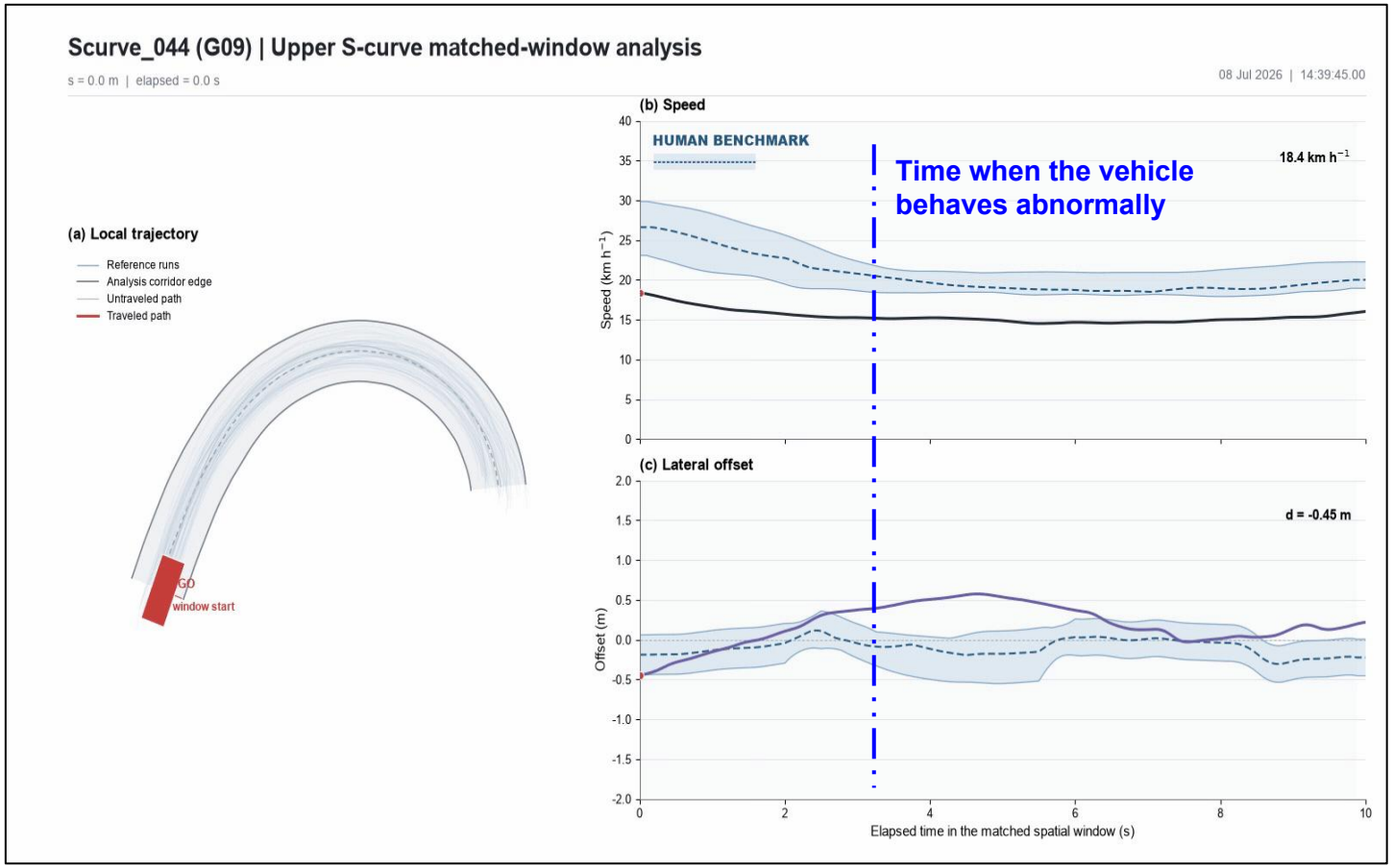
Results and Findings

Case Study (Point ②) : Non-Human-like Process

The AD drove at a lower speed but exhibited significant path deviation, straddling the solid center line.



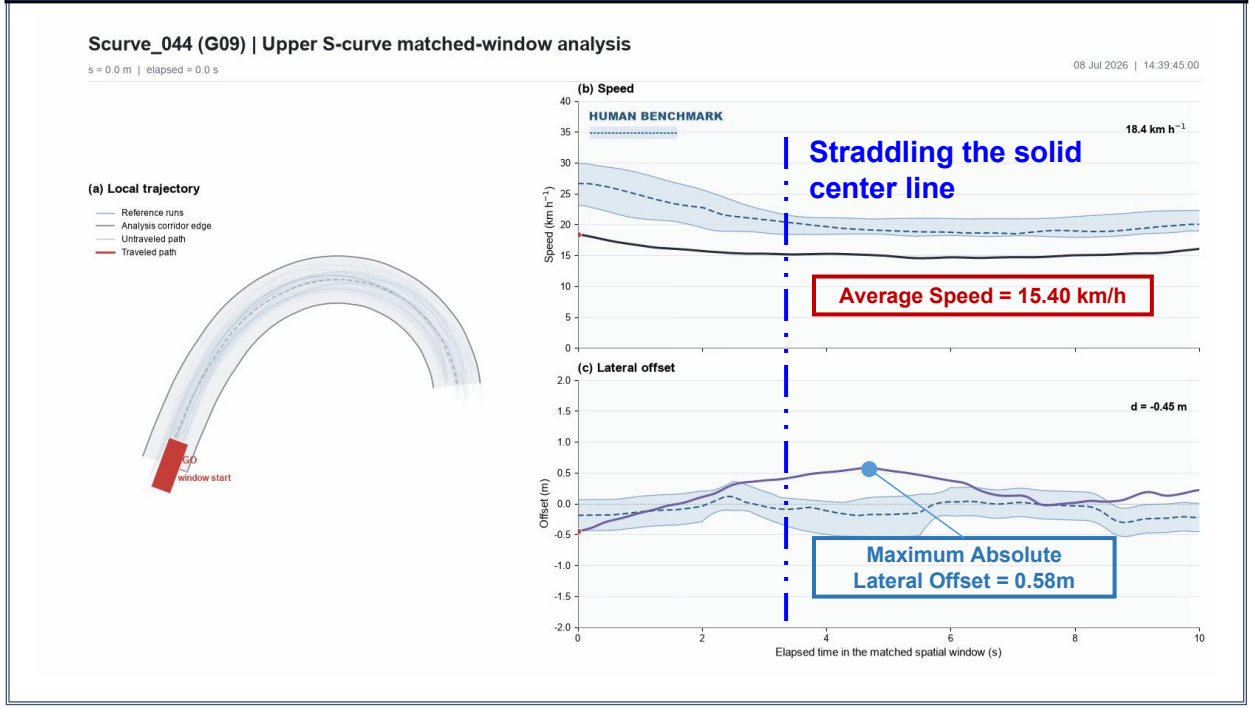
- **Where/when:** S-curve
- **What:** Unstable lane keeping
  (straddling the solid center line)
- **Why:** Improper planning
- **Intervention:** No



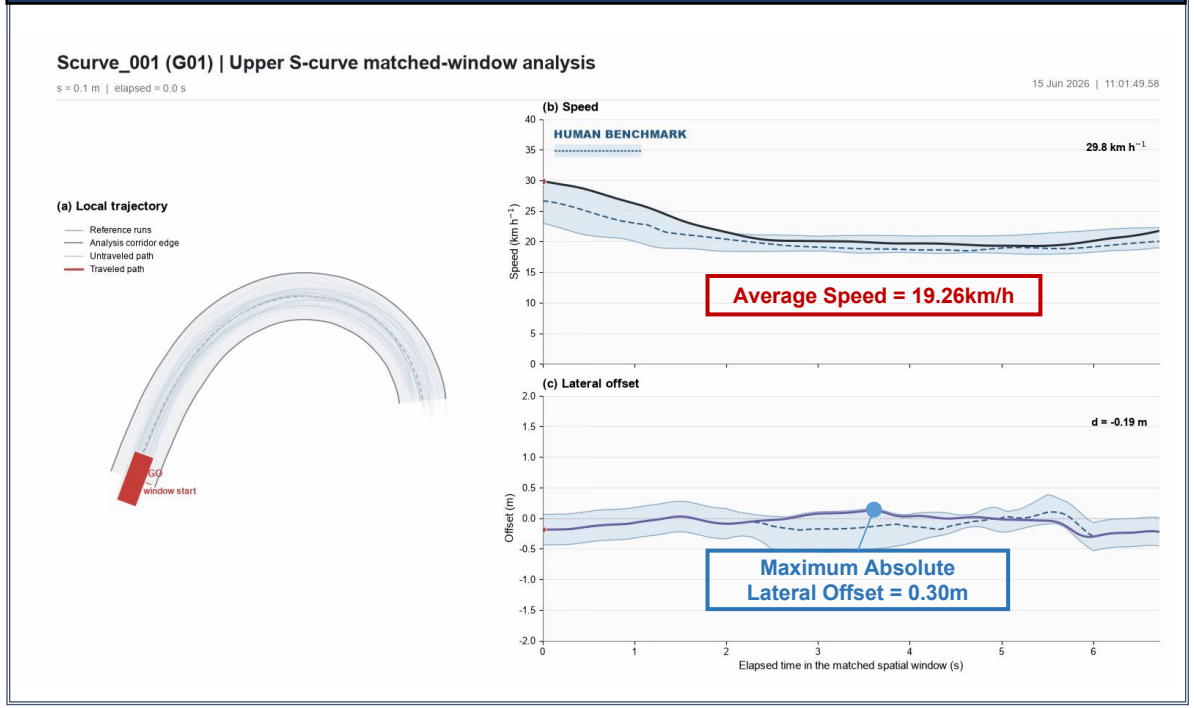
Results and Findings

Case Study (Point ②) : Non-Human-like Process

Non-Human-Like Driving



Human-Like Driving



	Average Speed (km/h)	Maximum Absolute Lateral Offset (m)
Non-Human-Like Driving	15.40	0.58
Human Benchmark	21.77 (17.90, 39.53)	0.40 (0, 0.40)

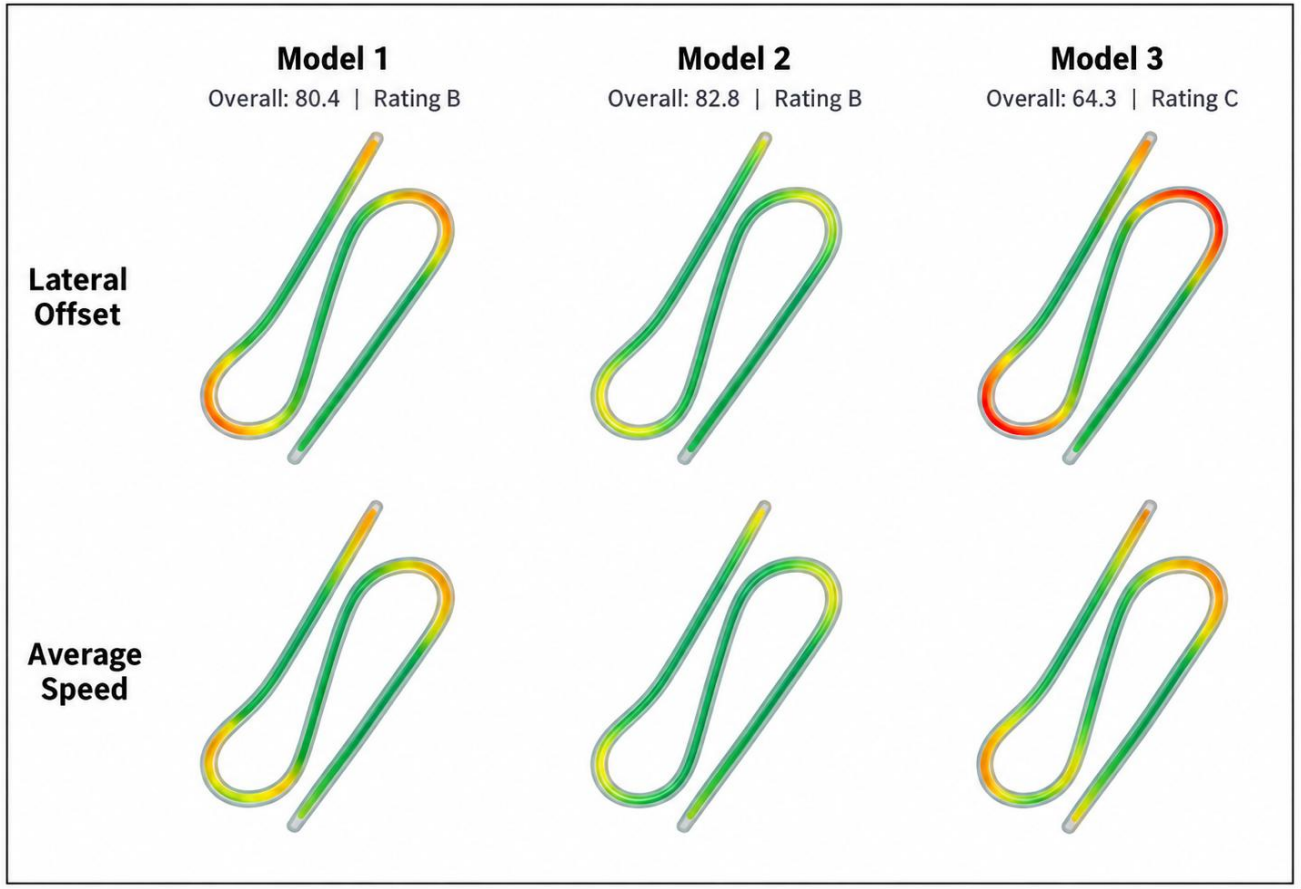
Non-human-like driving in S-curve: unexpected low speed with large lateral offset

Results and Findings

Human-Likeness Driving Performance of TEST A

Model	Score	Rating
Model 1	80.4	B
Model 2	82.8	B
Model 3	64.3	C

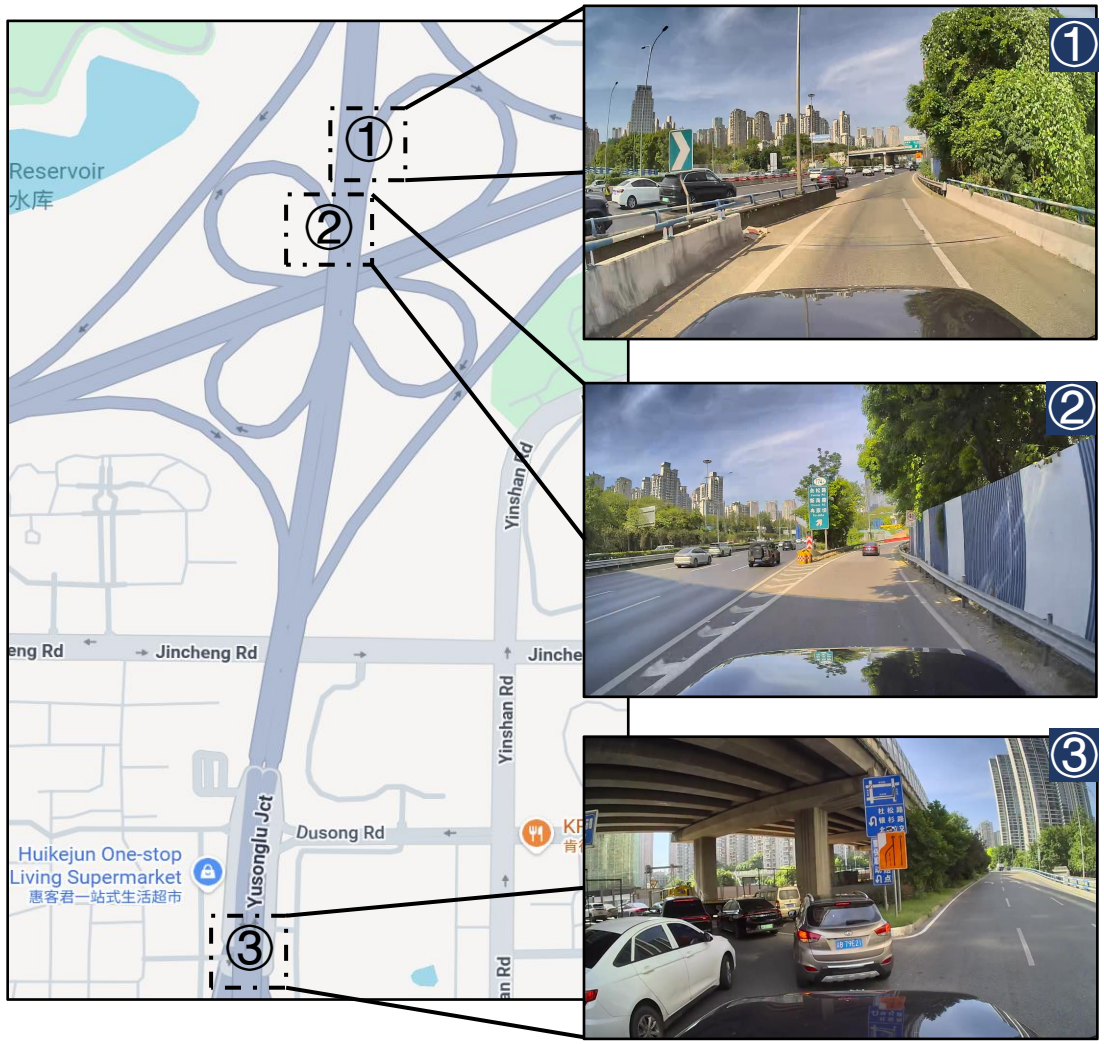
- All models show suboptimal performance in Lateral Offset and Average Speed, with substantial differences in Driving Rhythm relative to competent and careful human drivers.
- Model 3 shows less consistent lateral control and a more unstable Driving Rhythm, with a lower Lateral Offset score than the other two models.



Driving Rhythm, characterized by lateral offset and average speed in S-Curves, effectively captures behavioral differences between the tested vehicle models and competent and careful human drivers.

Results and Findings

TEST B: a high-density highway interchange featuring multiple merging, diverging, and weaving zones, creating challenging scenarios for interactive decision making of intelligent vehicles



Complex Road Topology

Closely spaced ramps, merges, and exits require accurate understanding of lane connectivity and route intent.

Frequent Vehicle Interaction

Vehicles on the mainline and ramps create dynamic gaps and negotiation demands during merging and diverging.

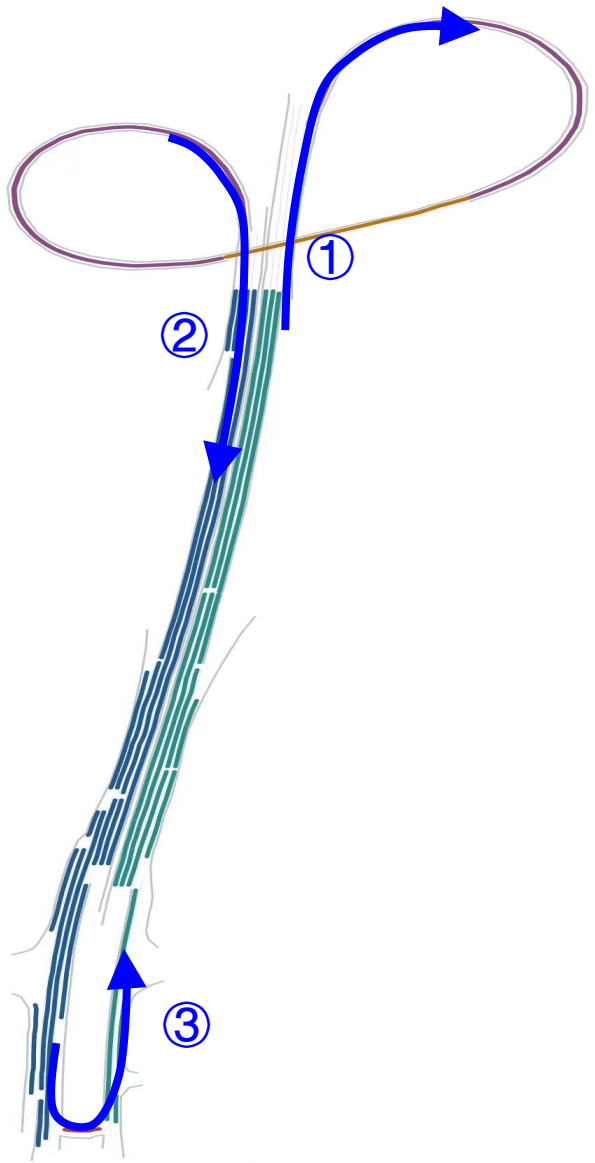
Limited Maneuvering Time

Limited distances between merging, lane-changing, and exiting sections leave little time for gap selection and maneuver execution.

Rapidly Changing Traffic Context

The surrounding traffic configuration changes continuously as vehicles enter, leave, and switch lanes, requiring repeated replanning.

Results and Findings



Total Mileage

289 km

2 vehicle models, **37** repeated rounds per model

Non-Human-Like Driving Events

318

1.10 events per kilometer

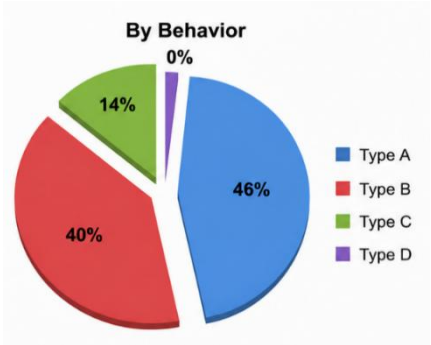
Trigger

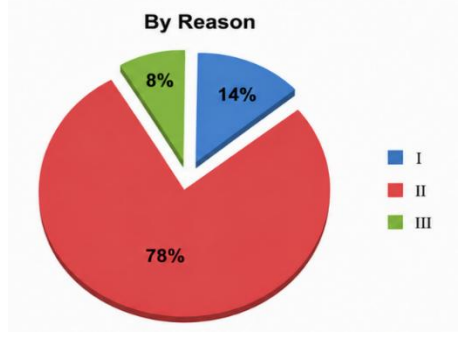
Takeovers

58

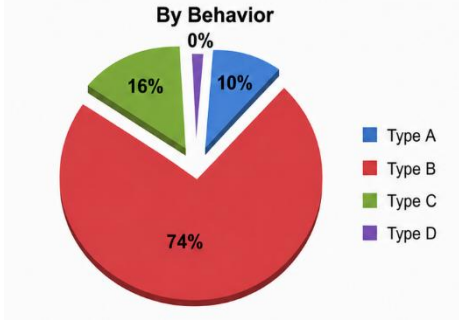
0.20 events per kilometer

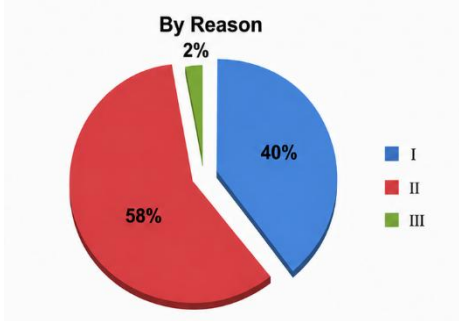
Non-Human-Like Driving Events





Takeovers





Abnormal Behavior Types

A. Unstable following or lane keeping

B. Inappropriate lane change/overtaking

C. Avoidance (obstacles, other vehicles, pedestrians etc.) anomalies

D. Start / Stop anomalies, etc.

Behavior Reason Annotation

I. Inaccurate Perception

II. Improper Planning

III. Incorrect Response

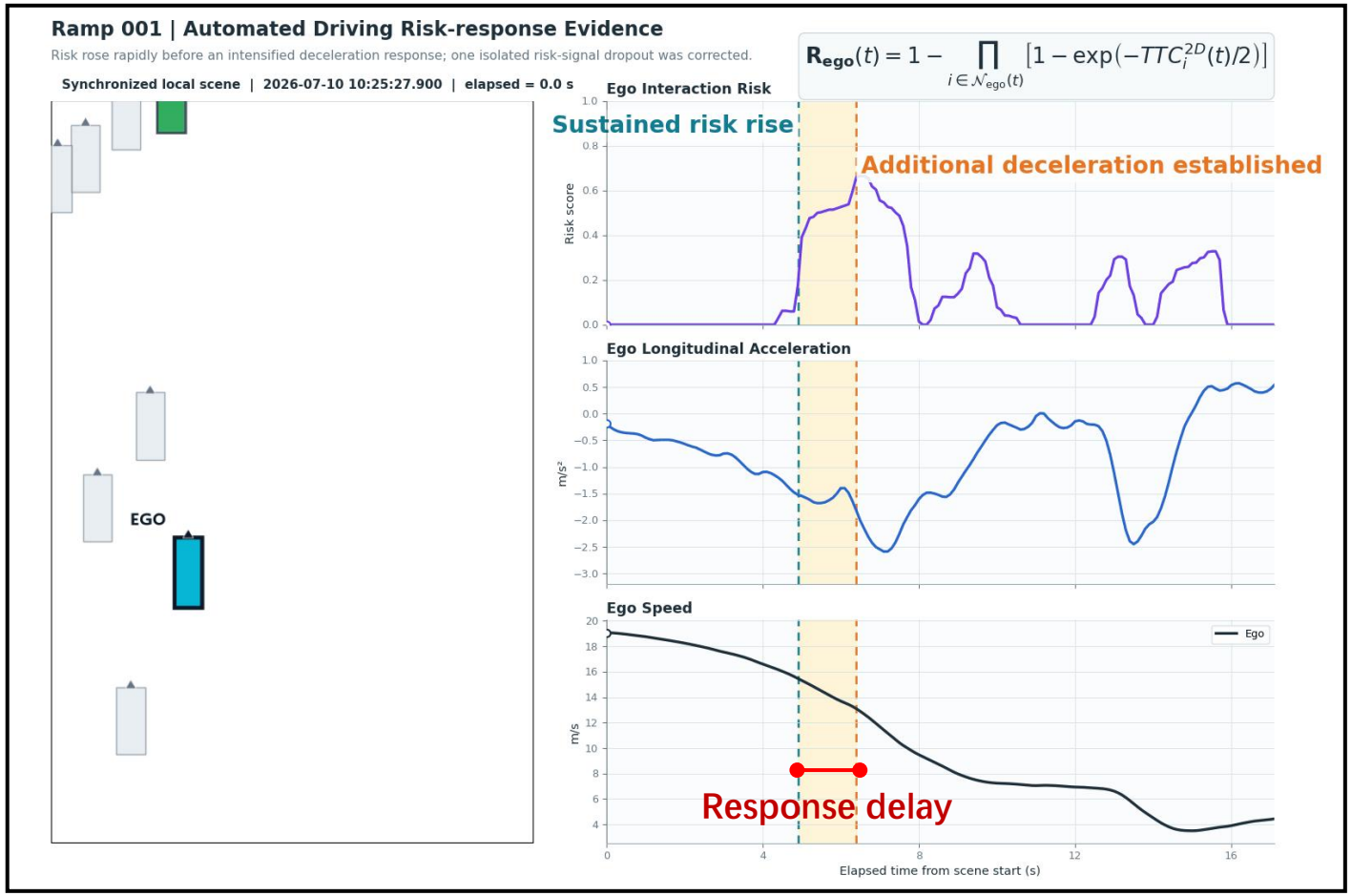
Results and Findings

Case Study (Point ①)： Non-Human-like Process

The AD showed a later respond to the target vehicle's merging intention, resulting in harder braking.



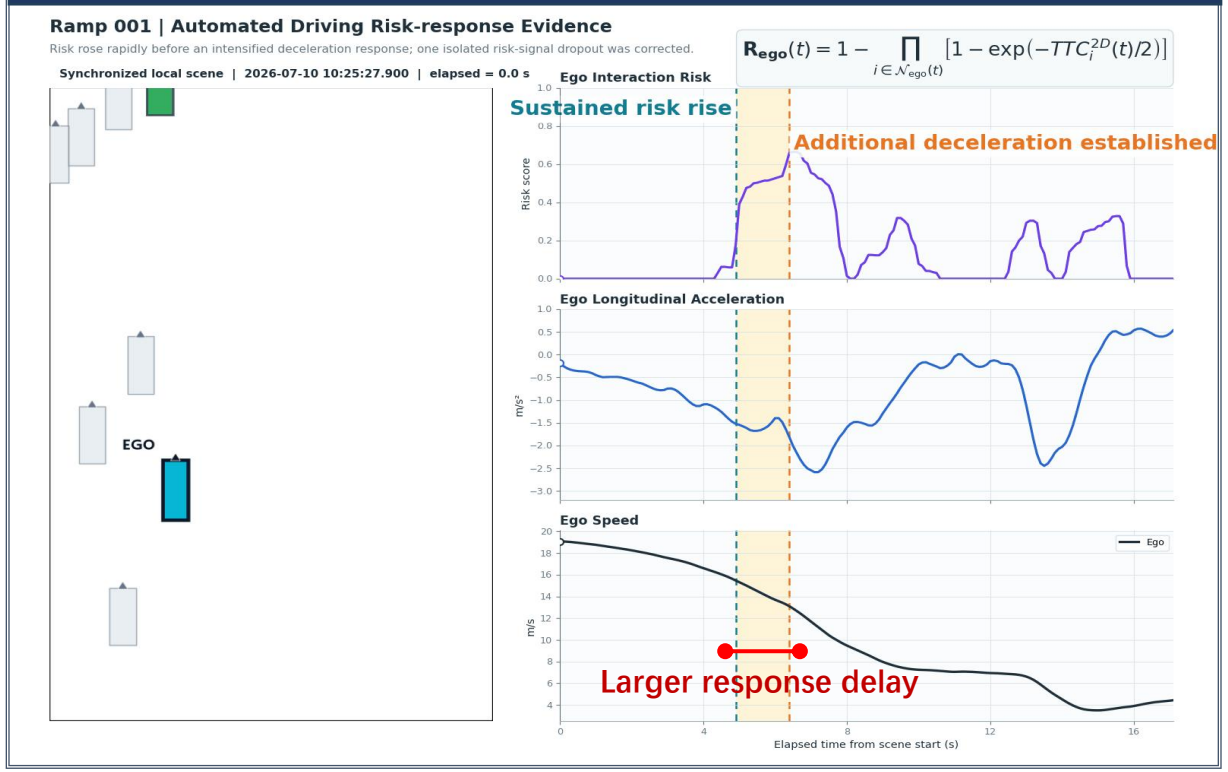
- **Where/when:** Weaving zone
- **What:** Avoidance of other vehicles anomalies (abrupt late braking)
- **Why:** Improper planning
- **Intervention:** No



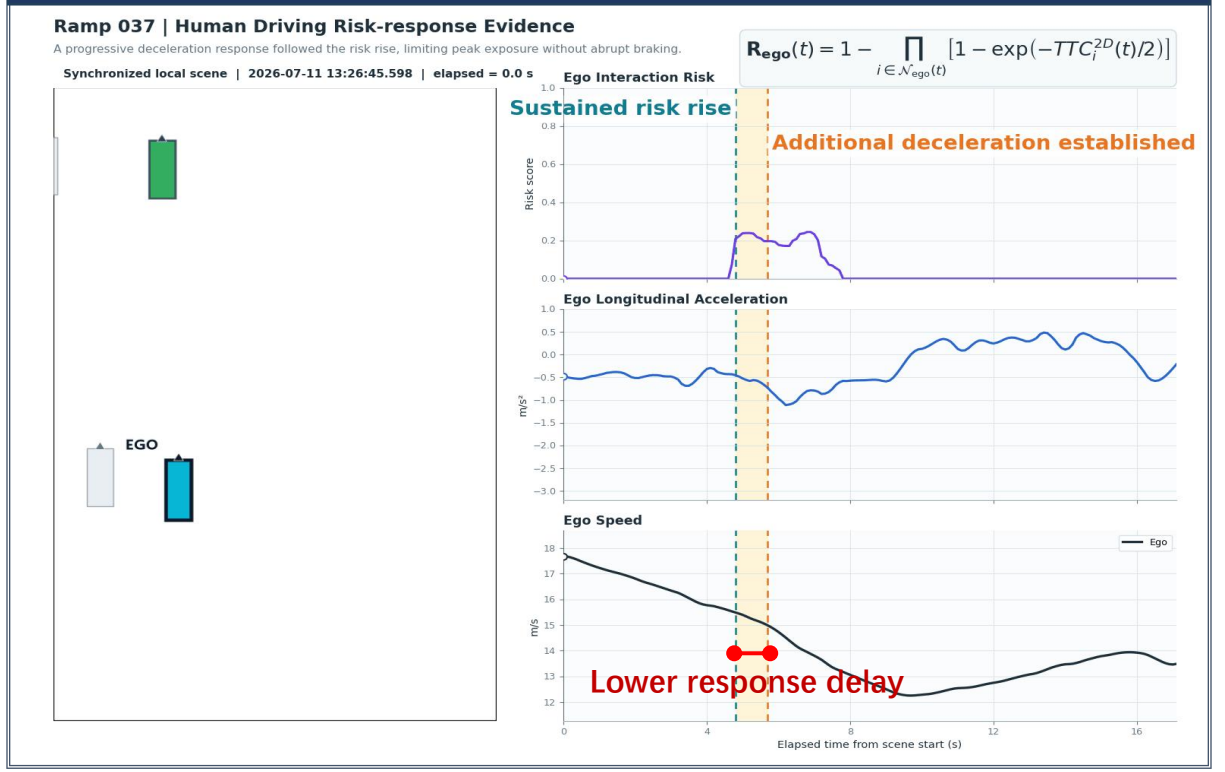
Results and Findings

Case Study (Point ①)： Non-Human-like Process

Non-Human-Like Driving



Human-Like Driving



	Peak Risk	Max Deceleration (m/s²)	Speed Loss (m/s)	Response Delay (s)
Non-Human-Like Driving	0.667	2.87	12.27	1.5
Human-Like Driving	0.245	1.25	3.49	0.9

Non-human-like driving in weaving zone: larger response delay with greater deceleration

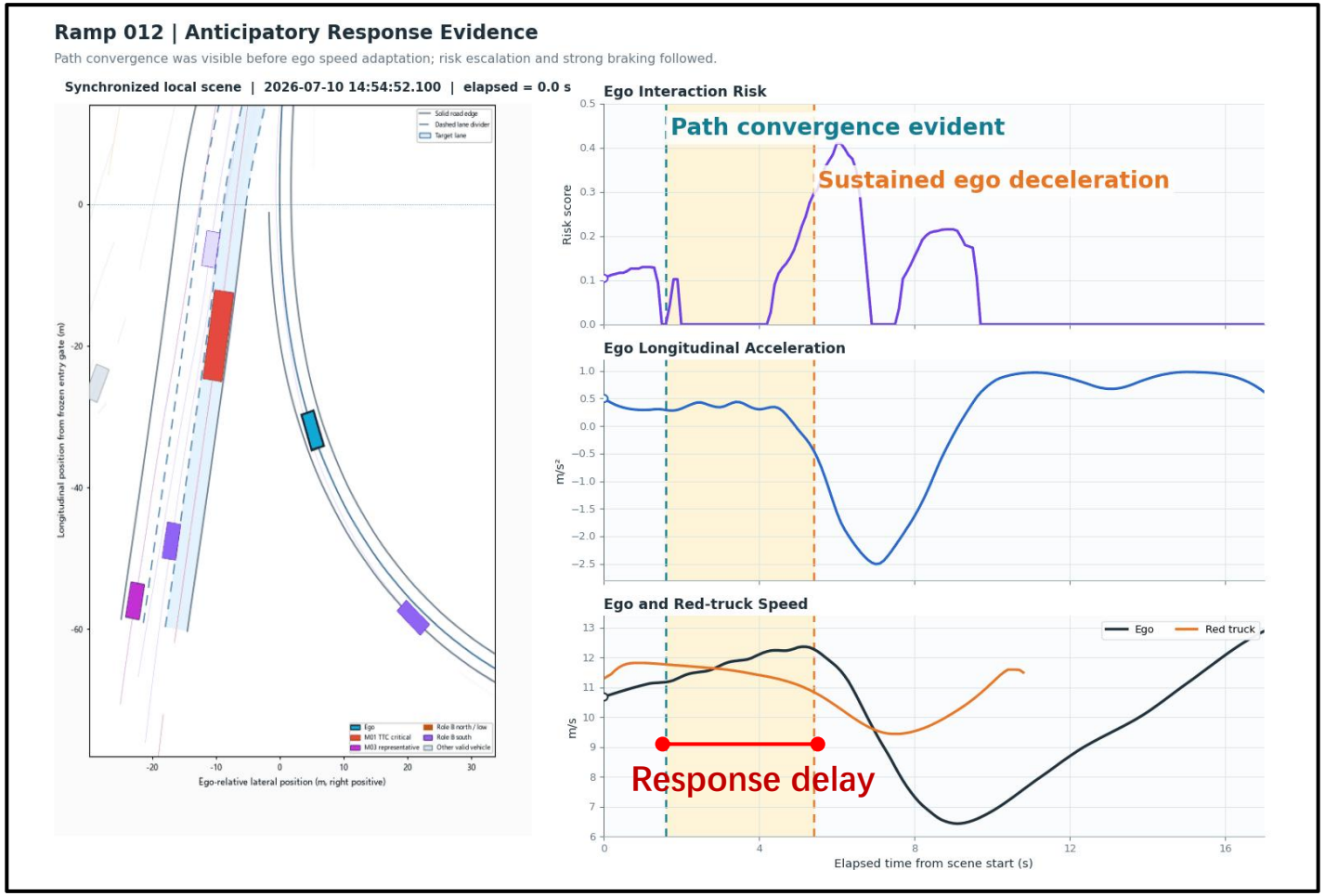
Results and Findings

Case Study (Point ②)： Non-Human-like Process

The AD showed a later respond to the target vehicle's diverging intention, resulting in harder braking.



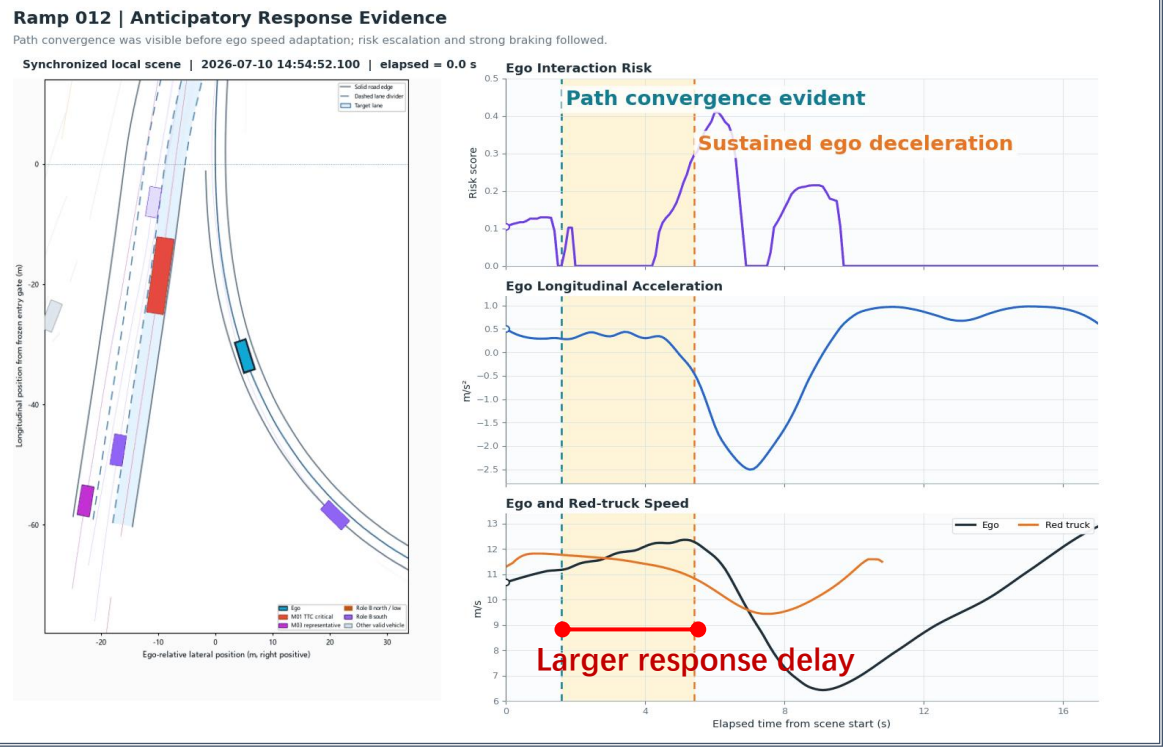
- **Where/when:** Merging zone
- **What:** Avoidance of other vehicles anomalies (abrupt Late Braking)
- **Why:** Improper Planning
- **Intervention:** No



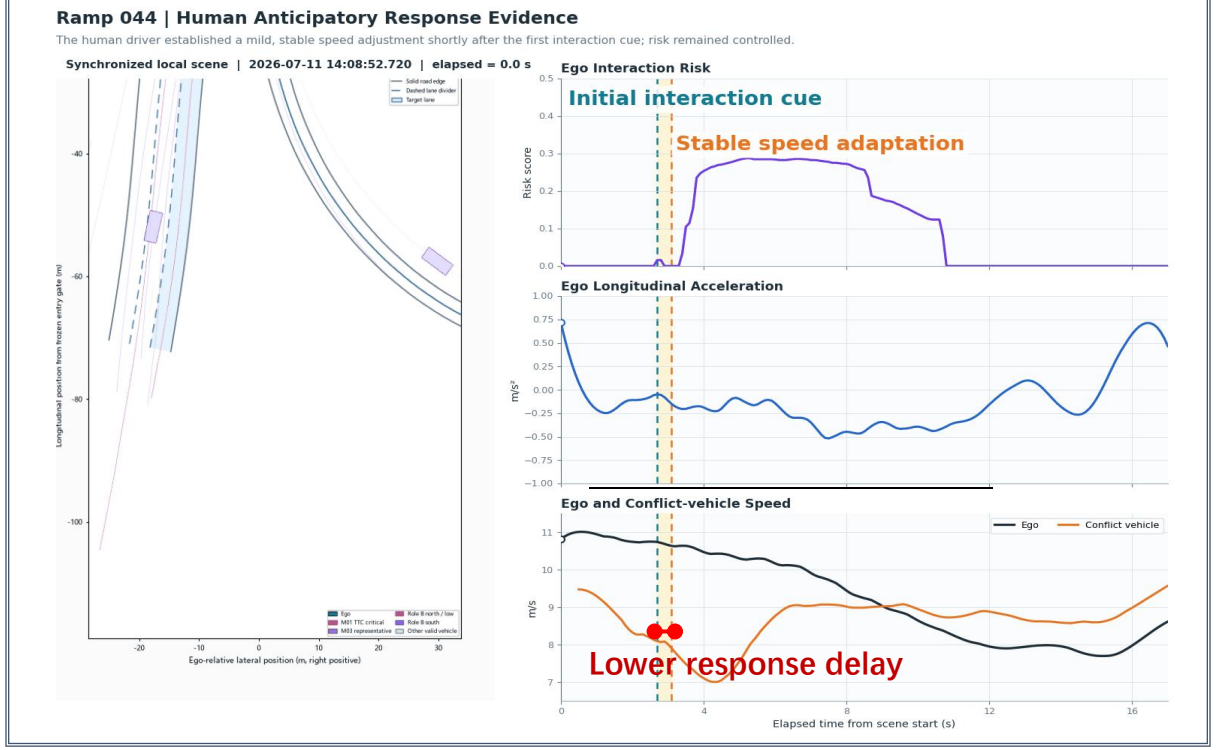
Results and Findings

Case Study (Point ②)： Non-Human-like Process

Non-Human-Like Driving



Human-Like Driving



	Peak Risk	Max Deceleration (m/s²)	Speed Loss (m/s)	Response Delay (s)
Non-Human-Like Driving	0.422	2.47	7.69	3.8
Human-Like Driving	0.286	0.52	6.48	0.4

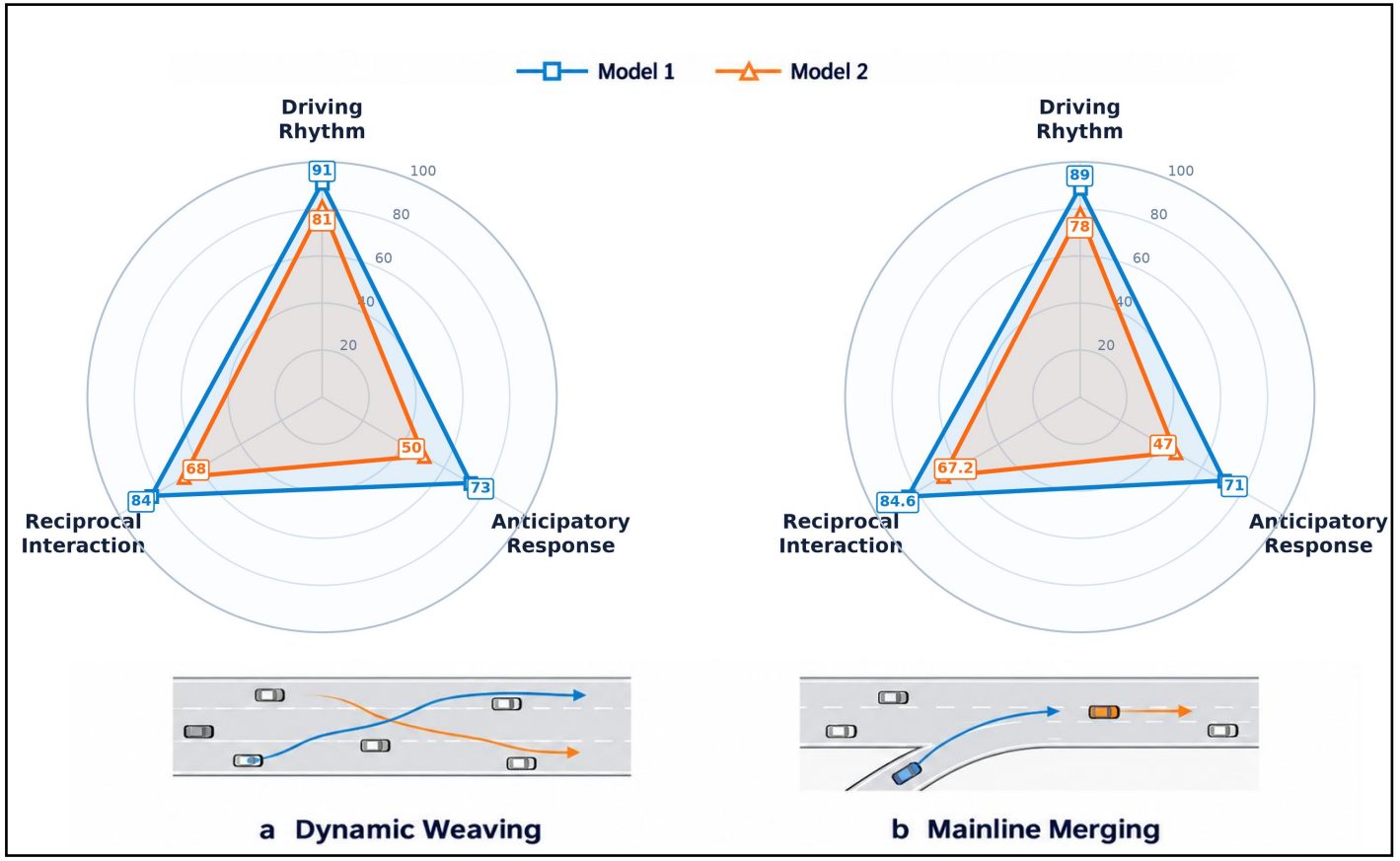
Non-human-like driving in weaving zone: larger response delay with greater deceleration

Results and Findings

Human-Like Driving Performance Assessment of TEST B

Model	Score	Rating
Model 1	82.1	B
Model 2	65.2	C

- Both models show substantially poorer Anticipatory response than competent and careful human drivers, with delayed responses when high-risk situations arise.
- Model 2 occasionally fails to anticipate other vehicles' maneuvers accurately, leading to overly aggressive responses that increase collision risk for both interacting vehicles.



Anticipatory response and reciprocal interaction in high-risk interactions reveal performance gaps between the tested vehicle models and competent and careful human drivers.

Outline

1.Background and Motivation

2.Methods and Experiments

3.Results and Findings

4.Future Prospects

Future Prospects

Expand the Scope and Scale of Testing



- *Broaden coverage across road types, traffic environments*
- *Develop a large-scale, cross-regional, multi-city intelligent driving dataset*
- *Establish a human-driving benchmark covering real-world traffic conditions*

Broader scenario coverage
Larger-scale data
Greater representativeness

Refine Assessment Framework



- *Refine the hierarchical index framework*
- *Optimize scenario-based assessment methods*
- *Improve the interpretability of assessment results*

More rigorous framework
More precise methods
More interpretable results

Enable Industry Applications



- *Identify issues and optimize with traceable solutions*
- *Clarify safety and liability, and align industry standards*
- *Reduce risks, enhance trust, and accelerate progress*

More unified standards
Clearer accountability
Greater human credibility

Future Prospects

A Globally Applicable Protocol for Human-Likeness Driving Assessment

Define Evaluation Framework → **Localize & Collect Data** → **Integrate into IVISTA**

Assessment Framework V1

NOW



- *Compare ADAS with the human baseline*
- *Define indicators and rating*

Global ADAS Dataset

JUL 2026 - 2027



- *Representative scenarios across multiple regions*
- *Human-driven + ADS trajectories and videos*

Protocol for IVISTA

2027+

Vehicle	Exam A	Exam B	Human-likeness	Certification
Vehicle 1	A	A	A	Certified
Vehicle 2	B	B	B	Certified
Vehicle 3	C	C	C	Certified
Vehicle 4	D	D	D	Certified
.....	E	E	E	Not Certified

- *Cross-country and cross-brand intelligent vehicle tests*
- *Repeatability validation and protocol refinement*

THANKS

Chen Wang, Ph.D.,

Professor of Intelligent Transportation Systems

Southeast University

15005172397

chen_david_wang@seu.edu.cn.