Proceedings of the 2021 5th CAA International Conference on Vehicular Control and Intelligence (CVCI), Tianjin, China, October 29-31, 2021

Intelligence Assessment of Automated Driving Systems Based on Driving Intelligence Quotient*

1st Yulei Wang
Department of Control Science and Engineering
Tongji University
Shanghai, China
wangyulei@jlu.edu.cn

2nd Meng Li
Department of Control Science and Engineering
Tongji University
Shanghai, China
2010460@tongji.edu.cn

3rd Yanjun Huang
Clean Energy Automotive Engineering Center
Tongji University
Shanghai, China
yanjun_huang@tongji.edu.cn

4th Hong Chen*
Department of Control Science and Engineering
Clean Energy Automotive Engineering Center
Tongji University, Shanghai, China
chenhong2019@tongji.edu.cn

2021 5th CAA International Conference on Vehicular Control and Intelligence (CVCI) | 978-1-6654-0847-9/21/$31.00 ©2021 IEEE | DOI: 10.1109/CVCI54083.2021.9661123

***Abstract*—When design, test and validate an intelligent agent, assessing its intelligence is essential. While autonomous vehicles (AVs) are deployed to a certain degree, it is still hard to assess their intelligence because it highly depends on tested scenarios but in real world tested scenarios are limited and far away from edges. Therefore, this paper attempts to propose an intelligence assessment approach for automated driving systems (ADS) based on behavior index (BI) and scenario complexity (SC). The main contributions of the scheme consist of three aspects: 1) proposing an intelligence assessment framework by following the idea of Turing test, 2) presenting a scenario bank for scenario complexity (SC) and behavior metrics for behavior index (BI), and 3) constructing a definition of driving intelligence quotient (DIQ) by the product of SC and BI. Finally, we present a lane-change scenario bank in Monte Carlo simulations to demonstrate the proposed assessment approach.**

***Index Terms*—Autonomous vehicle, automated driving system, intelligence assessment, scenario complexity, behavior index**

I. Introduction

Human-being is smart enough but incomplete to drive a car or conduct even more complicated tasks after a long time evolution. Instead, autonomous vehicles (AVs) use sensors, controllers, and actuators to replace human to perceive environment, analyze collected information and make driving decisions. But how to assess such replacement is good or what degree it is good? From the literature, note that National Institute of Standards and Technology (NIST) proposed a framework of autonomy of unmanned systems level, denoted by ALFUS, to facilitate characterizing and articulating autonomy for unmanned systems with qualitative metrics [5]. SAE (American society of automotive engineers) [13] defined the vehicle automation level from L0 (no automation) to L5 (fully automation) according to the functions and ODDs (operational design domains). However, this SAE criteria is not sufficient enough to distinguish two AVs in the same level. Therefore, we need to find a suitable or delicate index to assess the intelligence of different automated driving systems (ADSs).

This work was supported by the Natural Key Research and Development Program of China 2018YFB0105100. Corresponding author: Hong Chen

Considerable attention has been paid to the breakthrough of theoretical fundamentals and core technologies in AVs and their testing and assessment research. During the past three decades, there have been various military and civilian projects, such as DARPA Grand/Urban Challenge [7], European Land Robot Test Robot Competition [14] and Intelligent Vehicles Future Challenge (IVFC) [6] in China, dedicated to developing behavior assessment approaches in simulated urban environment and testing automated driving degree of AVs.

For testing AVs, Li et al. [10] declared a semantic diagram testing framework with potential benefits of scenario and functionality-based testing. Furthermore, Feng et al. [3] proposed the framework of naturalistic and adversarial driving environment (NADE) to enhance the existing life-like simulations to accelerate the test process. The efficiency of using NADE was improved by introducing prior knowledge of AVs and policy evaluation from deep reinforcement learning (DRL).

On scene and task complexity, Wang et al. [19] investigated a modeling and quantitative assessment for scene complexity, in which basic and auxiliary scene complexity models were established by using analytic hierarchy process (AHP) [12]. In comparison, Chen et al. [2] proposed an algorithm under different metrics to proving grounds of AV tests using a generative sample-based optimization approach and naturalistic logged driving data. Huang et al. [6] established a task-specific assessment model including metrics analysis, metrics preprocessing, weights calculation and task preference. 25 testing tasks in IVFC were labeled with three levels. Ma et al. [11] summarized the challenging problems in vehicle testing and evaluation by providing a hierarchical architecture with target, criteria, calculation and result layers.

To evaluate the behavior and performance of AVs, Sun et al. [18] proposed a quantitative assessment method based on the chaos theory. The Lyapunov exponent of deviation time series was calculated and presented as the metric to evaluate tracking performance. In [20], an assessment system with multi-metrics was designed where the entropy-cost function method was applied to calculate their weights.

Huang et al. [4] proposed an integrated architecture for AVs' evaluation, in which the action amount of driving process was considered as the index of such evaluation, and the least action calculated from the best AV was employed as a standard to quantify behavior of other AVs. Chen et al. [1] proposed an adaptive assessment scheme for AVs in adversarial environment generated by deep reinforcement learning (DRL). Simulation results show that adversarial scenarios by learning can significantly degrade performance of tested vehicles.

In addition, Turing test is well-established and utilized to assess software candidates smart or not. In Turing test, a human evaluator asks the software some questions from the prepared question bank. Then, the intelligence of a software is given as a percentage indicating how much the software is as smart as a human. Obviously, an AV system is quite different from a piece of software. But the idea of Turing tests can be extended to calculate the driving ability of an AV.

Motivated by the literature, this paper makes some modifications before introducing Turing test and proposes a novel concept, driving intelligence quotient (DIQ), to assess ADSs more precisely. Specifically, a scenario bank is constructed to test vehicle behavior and performance instead of Q&A. Two models are established to replace a human evaluator so as to assess scenario complexity (SC) and behavior index (BI). Intuitively, we define DIQ of an AV as the product of SC and BI. To show the effectiveness of the proposed approach, a lane-change case is studied, in which SC, BI and DIQ are assessed in a quantitative way.

The rest of the paper is given as follows. In Section II, the intelligence assessment framework followed by Turing test is proposed by modeling SC and BI. Details of the scenario bank and behavior metrics are described in Section III. Section IV develops two specific models for SC with four test cases and BI with four typical metrics in lane-change maneuver. Results of Monte Carlo simulations for four candidates using DRL are presented. Section V draws the conclusion.

II. Framework of Intelligence Assessment

Let's first discuss the intelligence of AVs. As shown in Fig. 1, Famous Turing test is utilized to examine a software candidate by means of asking questions. If the candidate passes the test according to a percentage, it means the software is somehow smart. So, in this paper, we introduce intelligence to differentiate AVs. Obviously, an ADS is very different from a piece of software. But the idea of Turing test can be extended to calculate intelligence of AVs.

In Turing test, a human judge is used to ask the software several questions from the prepared question bank. Then, the intelligence of a software can be denoted by a percentage, indicating how much the software is as smart as a human. However, for an AV, it is impossible to directly apply the Turing test because Q&A is making no sense to a vehicle. Consequently, we make some modifications before applying the Turing test. As shown in Fig. 2, a scenario bank is first built to test vehicle behaviors instead of the Q&A procedure. Furthermore, we propose two models to replace the human evaluator so as to evaluate SC and BI, which is closely

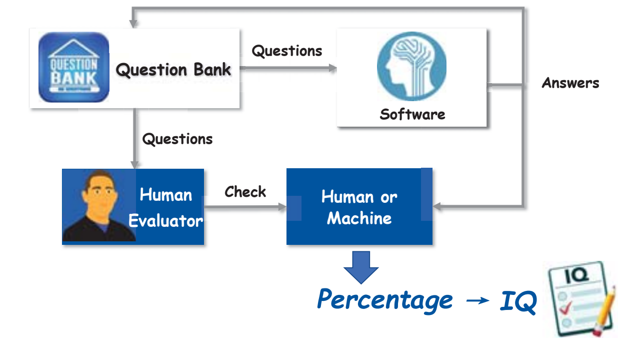

Fig. 1. A schematic diagram of Turing test for softwares

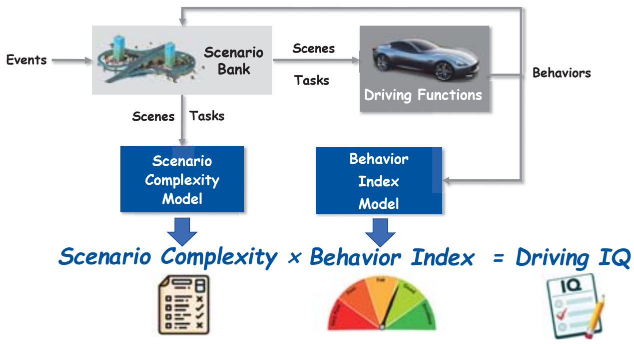

Fig. 2. A schematic diagram of intelligence assessment for AVs

related to the intelligence of AVs. Hence, we define driving intelligence quotient (DIQ) of AVs as the product of SC and BI, i.e.,

$$IQ = SC \times BI, \tag{1}$$

where SC and BI are further modeled in the next section.

III. Scenario Bank and Behavior Metrics

A. Scenario Bank

A scenario bank is utilized to describe specific operational design domains (ODDs) and functions of dynamic driving tasks (DDTs) that include the capability of identifying object and event detection and response (OEDR). In AV testing, the scenario bank is pre-built either in simulation environment, HIL, or real-world environment, where AV systems are tested with a set of given tasks.

In theorem, a scenario bank can be divided into scenes and tasks. Referring to [15], a scene taxonomy for a scenario is normally hierarchical and includes the following top-level categories:

- Physical Infrastructure
- Operational Constraints
- Objects
- Connectivity
- Environmental Conditions
- Special Zones

Some of challenges associated with scene elements include their variability, risk degree and frequency of occurrences. Hence, the scene elements are divided into basic and auxiliary elements for a scene, in which elements of basic scenes include physical infrastructure, operational constraints and

TABLE I
SUMMARY OF VEHICLE BEHAVIOR METRIC CATEGORIES

Literature	Human Interface	Mission Complete	Accuracy	Safety	Comfort	Efficiency	Rationality	Learnability
Huang [5]	✓							
Sun et al. [17]		✓				✓		
Sun et al. [16]		✓						
Huang et al. [6]		✓	✓	✓				
Li et al. [9]	✓							
Sun et al. [18]			✓					
Zhao et al. [20]		✓	✓	✓	✓			
Li et al. [10]		✓		✓	✓	✓		
Huang et al. [4]			✓	✓	✓		✓	
Ma et al. [11]		✓	✓		✓			
Chen et al. [1]		✓	✓	✓				✓

objects, and elements of auxiliary scenes refer to connectivity, environmental conditions and special zones.

Note that each top-level element mentioned above can be further divided into many sub-classes. However, this paper will not go to the details because of the space limit. Readers are recommended to refer [15] for more details.

Similarly, the tasks consist of: [15]

- Tactical Maneuvers
- OEDRs

where tactical maneuvers mainly have parking, lane change, car following, lane maintenance, and navigate Intersection, etc, OEDRs are identified with respect to vehicle, infrastructure, vulnerable road user, animals, etc. Details of sub-class elements for tasks can be referred to [15].

B. Behavior Metrics

In order to assess how much an AV is as smart as human, we need to identify comprehensive evaluation metrics from the literature and present a representative subset of the reviewed work. As shown in Table I for an overview, the metrics for behavior index include human interface, mission completion, safety, accuracy, comfort, efficiency, rationability and learnability. The statements of these metrics are as follows.

- Human Interface [5]: The metric addresses the activity by which human operators engage with AVs to achieve mission goals. Note that frequency, duration, workload and operator to AV ratio are the factors according to amount and relative importance.
- Mission Completion [10]: The metric describes whether a vehicle has successfully fulfilled each desired mission within the appropriate temporal-spatial ranges, and without violating traffic laws.
- Safety [10]: The metric is mainly calculated according to the risk or danger level that a vehicle has encountered during the travel in the assigned traffic scenario. For example, distances between the testing vehicle and other road users (vehicle, bicycle, pedestrian, obstacle, stop line) can characterize the safety level. The closer two objects are, the smaller the safety score is.
- Accuracy [6]: The metric reflects the errors between actual values and desired values such as position estimation errors, path-following errors, road detection errors, etc.
- Comfort [10]: The metric measures the physical output of a vehicle to indicate feelings of human passengers, where longitudinal ac/deceleration rates and jerks are commonly used in vehicle tests.
- Efficiency [17]: The metric is utilized to describe AVs how to accomplish given tasks. Normally, the time and energy consumption for completing a task is less, the efficiency score is higher.
- Rationality [4]: Rationality reflects the psychological expectation of human drivers, and indicates the difference between real traffic behaviors and expected traffic behaviors. For instance, when there are other road users, a smart AV should pursue consistency of driving characteristics between the ego vehicle and surrounding vehicles, resulting in steady traffic flow.
- Learnability [1]: The metric is used to describe the time evolution principle of the action performance of AVs in dynamic environment. If an AV operates better performance with time, the learning score of the AV is positive. Otherwise, the score is negative.

IV. MODELS OF SCENARIO COMPLEXITY AND BEHAVIOR INDEX FOR LANE-CHANGE MANEUVERS

In the section, we introduce a lane-change scenario bank to show how to calculate SC and BI by the proposed framework.

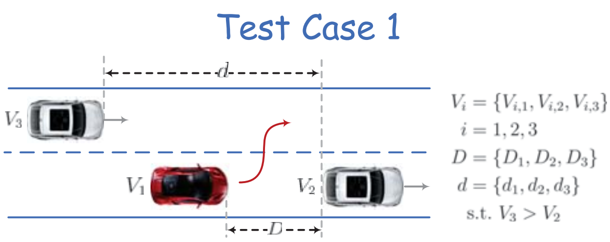

Fig. 3. Schematic diagram of Test Case 1, where Monte Carlo simulations process with a random initialization from $V_1 = \{20, 25, 30\}$ m/s, $V_2 = V_1 - \{5, 8, 12\}$ m/s, $V_3 = V_1 - \{1, 10\}$ m/s, $D = \{35, 50, 70\}$ m and $d = \{45, 60, 90\}$ m.

A. Scenario Complexity Model

The lane-change scenario bank consists of four test cases, denoted by TC-$\{1, \ldots, 4\}$. Schematic diagrams of four test cases are shown in Figs. 3-6. Test Case 1 considers a scenario of lane change where the front and left rear of the ego vehicle

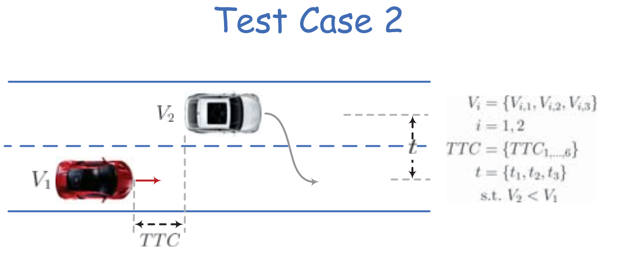

Fig. 4. Schematic diagram of Test Case 2, where Monte Carlo simulations process with a random initialization from $V_1 = \{20, 22, 25\}$, $V_2 = V_1 - \{3, 5, 8\}$ and $TTC = \{3.75, 6, 7, 8, 10, 13.3\}$ s. The object vehicle is going to change lane.

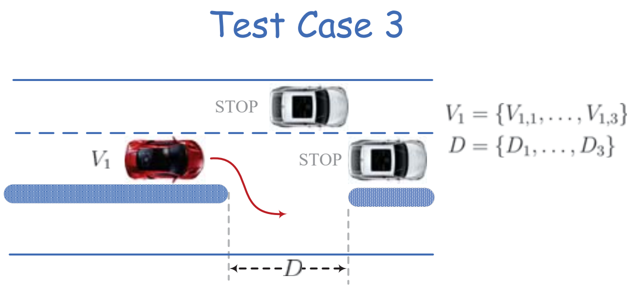

Fig. 5. Schematic diagram of Test Case 3, where Monte Carlo simulations process with a random initialization from $V_1 = \{20, 22, 25\}$ and $D = \{30, 40, 50\}$ m. The object vehicles are stopping.

have target vehicles. Test Case 2 addresses a scenario of lane change where the target vehicle changes to the current lane. Test Case 3 investigates a scenario of lane change into an exit ramp with a width D. Finally, Test Case 4 takes an on-ramp with a distance D.

To model SC, the speed of ego vehicle, TTC-1 (time-to-collision to the front vehicle), TTC-2 (time-to-collision to the vehicle in the target lane) and the flag whether the target vehicle changes lane are considered. The empirical formulas of scenario complexity are as follows:

$$C_V = \begin{cases} 0 & V < 0 \\ \frac{1}{5}V & V \in [0, 50] \\ 10 & V > 50 \end{cases} \tag{2}$$

$$C_{TTC} = \begin{cases} 0 & TTC > 3 \\ \frac{10}{3}(3 - TTC) & TTC \in [0, 3] \\ 10 & TTC < 0 \end{cases} \tag{3}$$

$$C_{LC} = \begin{cases} 0 & \text{Lane Change = N} \\ 10 & \text{Lane Change = Y} \end{cases} \tag{4}$$

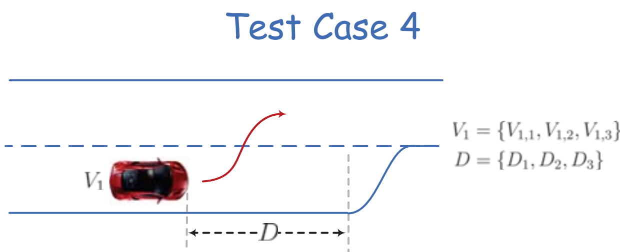

Fig. 6. Schematic diagram of Test Case 4, where Monte Carlo simulations process with a random initialization from $V_1 = \{20, 22, 25\}$ and $D = \{30, 50, 70\}$ m.

Based on these equations, the SC model is established as

$$SC = \omega_1 C_V + \omega_2 C_{TTC1} + \omega_3 C_{TTC2} + \omega_4 C_{LC} \tag{5}$$

where weights of complexity ω_i are determined by AHP [12] and SC scores for the four test cases are list in Table II.

TABLE II
SCORES OF SCENARIO COMPLEXITY FOR FOUR TEST CASES

AHP		-	C_V	C_{TTC1}	C_{TTC2}	C_{LC}	Scores
ω_1	0.15	TC-1	5.00	0.00	0.00	0.00	0.75
ω_2	0.30	TC-2	5.00	0.00	0.00	10.0	3.75
ω_3	0.25	TC-3	4.50	2.30	6.30	0.00	2.96
ω_4	0.30	TC-4	4.50	6.00	0.00	0.00	2.48

B. Behavior Index Model

To validate behavior and performance of AVs, the tested system candidates should be defined. As shown in Fig. 7, autonomous lane change systems based on DRL are introduced to be tested candidates. Note that environment of DRL algorithms is set at Highway-Env [8]. The full connected neural network is built with 25 vehicle states, 5 vehicle responses and two 256 hidden layers. The reward function is given by

$$R(s, a) = \omega_a \frac{V_x - V_{x,\min}}{V_{x,\max} - V_{x,\min}} - \omega_b \cdot \text{collision} \tag{6}$$

and the loss function is set as mean square error (MSE).

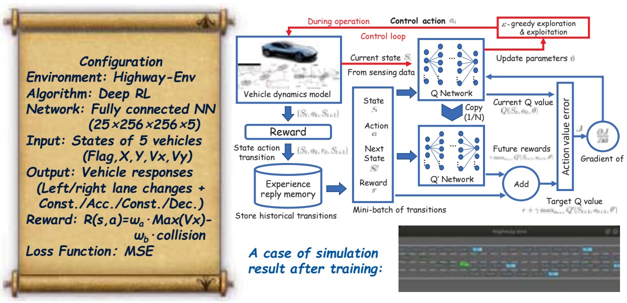

Fig. 7. Autonomous lane change algorithm: DRL.

To explain the rationality or effectiveness of the proposed assessment approach, we propose four candidates to calculate their DIQs. Note that Candidates 1 and 2 (NN$-$1 and NN$-$2) are trained 7 and 10 thousand times in open-source Highway-Env dataset, respectively. Candidates 3 and 4 (NN$-$3 and NN$-$4) are trained another 5 thousand times in the scenario bank on the basis of Candidates 1 and 2, respectively.

Intuitively, the behavior indexes of safety P_{safe}, mission completion $P_{mission}$, rationality P_{ration} and learnability P_{learn} are denoted as

$$P_{safe} = 10 \times \frac{n - n_{col}}{n} \tag{7}$$

$$P_{mission} = 10 \times \frac{n_{lc}}{N - n_{col}} \tag{8}$$

$$P_{ration} = 10 \times \frac{n_{lc3}}{n_{lc}} \tag{9}$$

$$P_{learn} = 10 \times \frac{R - R_{\min}}{R_{\max} - R_{\min}} \tag{10}$$

where n is the number of the total test, n_{col} is the number of collision, n_{lc} is the number of lane change, n_{lc3} is the number satisfying lane change less than 3, and R is the average reward point for each candidate.

Based on the equations (7)-(10), we propose the behavior index model as

$$BI = \rho_1 P_{safe} + \rho_2 P_{miss} + \rho_3 P_{rati} + \rho_4 P_{lear} \tag{11}$$

where the weight ρ_i is calculated by AHP and the index elements in (11) are computed by collecting data from the four candidates under four test cases.

The weight of BI is calculated by $\rho = [0.3, 0.3, 0.2, 0.2]$. The simulation results are shown in Figs. 8-11 and summarized in Tables III-VI. Note that P_{safe} indexes increase after training, but the learning algorithms cannot ensure to improve the indexes of $P_{mission}$, P_{ration} and P_{learn}. It is reasonable because the reward function in DRL only considers the speed maximum and collision avoidance.

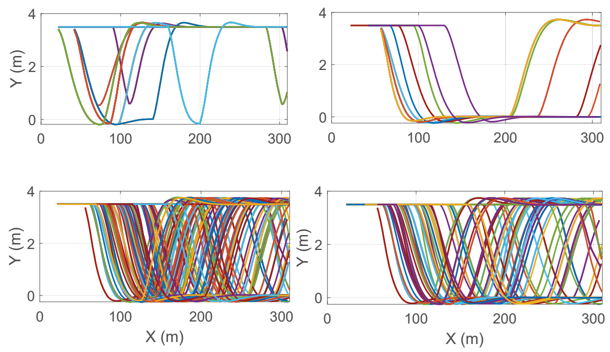

Fig. 8. Simulation results of Test Case 1.

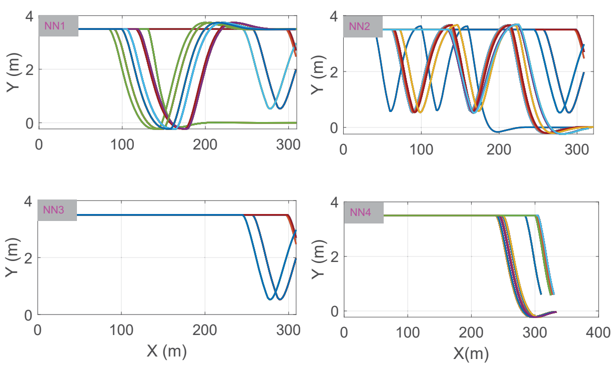

Fig. 9. Simulation results of Test Case 2.

TABLE III
SCORES OF BI UNDER TEST CASE 1

-	P_{safe}	$P_{mission}$	P_{ration}	P_{learn}	Scores
NN-1	4.10	1.11	7.02	0.81	4.52
NN-2	4.05	7.50	3.33	4.73	5.49
NN-3	9.85	9.41	9.21	6.63	9.00
NN-4	9.90	10.0	8.62	6.90	8.92

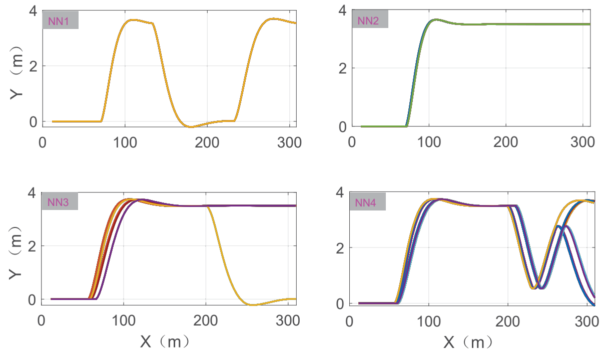

Fig. 10. Simulation results of Test Case 3.

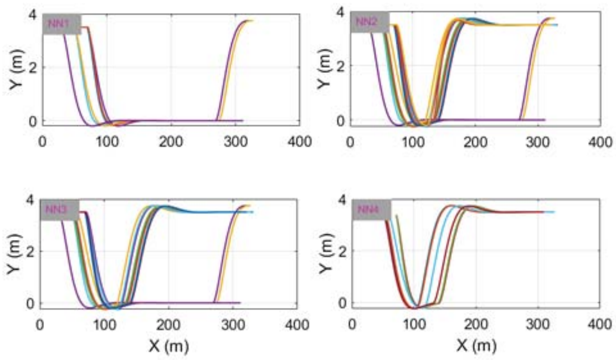

Fig. 11. Simulation results of Test Case 4.

TABLE IV
SCORES OF BI UNDER TEST CASE 2

-	P_{safe}	$P_{mission}$	P_{ration}	P_{learn}	Scores
NN-1	3.50	8.07	7.14	10.0	7.48
NN-2	3.00	5.09	0.00	3.60	4.32
NN-3	10.0	9.87	6.04	8.06	6.54
NN-4	10.0	9.63	10.0	7.83	7.99

TABLE V
SCORES OF BI UNDER TEST CASE 3

-	P_{safe}	$P_{mission}$	P_{ration}	P_{learn}	Scores
NN-1	1.20	3.33	0.00	1.55	3.67
NN-2	2.00	0.00	10.0	0.00	5.60
NN-3	10.0	9.92	6.85	7.56	8.88
NN-4	10.0	9.84	0.00	7.44	7.49

TABLE VI
SCORES OF BI UNDER TEST CASE 4

-	P_{safe}	$P_{mission}$	P_{ration}	P_{learn}	Scores
NN-1	10.0	4.81	10.0	1.63	8.33
NN-2	2.55	2.92	10.0	2.02	6.17
NN-3	10.0	8.27	10.0	5.85	9.17
NN-4	10.0	9.35	6.65	6.67	8.66

TABLE VII
SCORES OF DIQ FOR FOUR CANDIDATES UNDER FOUR CASES

-	DIQ-1	DIQ-2	DIQ-3	DIQ-4	Total DIQ
NN-1	2.35	25.88	4.94	16.78	49.95
NN-2	3.81	11.80	7.70	10.03	33.34
NN-3	6.71	32.93	26.22	21.45	87.31
NN-4	6.81	35.46	22.02	21.00	85.29

Fig. 12. DIQ rank of four candidates.

Based on (1), DIQ scores of the four candidates under four test cases are computed and shown in Table VII. The rank is vividly shown in Fig. 12. The results indicate the conclusions that Candidate 3 wins the best ADS because post-training improves the DIQ of candidates, which demonstrates the effectiveness of the proposed intelligence assessment approach.

V. CONCLUSIONS

AV is hot research area all over the world such that a more comprehensive assessment approach should be proposed to keep pace with the development of technology research. That is why in this talk we have shown what our team is doing now in this topic, including how we extend Turing test to the vehicle area and introduce the DIQ model that encourages AVs to challenge edge scenarios by high scenario complexity. But, regarding DIQ testing and assessment, we just began and there is a long way to go before we have a more generic model or conclusion.

ACKNOWLEDGMENT

The authors would like to thank Puhang Xu, the postgraduate of Tongji University, for his technical support on Monte Carlo simulations.

REFERENCES

[1] B. Chen, X. Chen, Q. Wu, and L. Li, “Adversarial evaluation of autonomous vehicles in lane-change scenarios,” *IEEE Trans. Intell. Transp. Syst.*, 2021, in Press.

[2] R. Chen, M. Arief, W. Zhang, and D. Zhao, “How to evaluate proving grounds for self-driving? A quantitative approach,” *IEEE Trans. Intell. Transp. Syst.*, 2020, Early Access, DOI: 10.1109/TITS.2020.2991757.

[3] S. Feng, X. Yan, H. Sun, Y. Feng, and H. X. Liu, “Intelligent driving intelligence test for autonomous vehicles with naturalistic and adversarial environment,” *Nat. Commun.*, vol. 12, no. 787, pp. 1–14, 2021.

[4] H. Huang, X. Zheng, Y. Yang, J. Liu, W. Liu, and W. J., “An integrated architecture for intelligence evaluation of automated vehicles,” *Accid. Anal. Prev.*, vol. 145, no. 105681, pp. 1–12, 2020.

[5] H. M. Huang, “Autonomy levels for unmanned systems (ALFUS) framework, Volume i: Terminology,” in *Proc. of the 2007 PerMIS*, Gaithersburg, USA, 2007, pp. 48–53.

[6] W. Huang, D. Wen, J. Geng, and N. N. Zheng, “Task-specific performance evaluation of UGVs: Case studies at the IVFC,” *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 5, pp. 1969–1979, 2014.

[7] E. Krotkov, S. Fish, L. Jackel, B. McBride, P. M., and J. Pippine, “The DARPA PerceptOR evaluation experiments,” *Auton. Robots*, vol. 22, no. 1, pp. 19–35, 2006.

[8] E. Leurent, “An environment for autonomous driving decision-making,” https://github.com/eleurent/highway-env, 2018.

[9] J. Li, Z. Tang, Y. Fang, and Y. Tan, “Autonomy evaluation methods for the autonomous ground vehicles,” in *Proc. of the 2015 ICMMCCE*, Xi’an, China, 2015, pp. 1347–1356.

[10] L. Li, W. Huang, Y. Liu, N. N. Zheng, and F. Y. Wang, “Intelligence testing for autonomous vehicles: A new approach,” *IEEE Trans. Intell. Transp. Syst.*, vol. 1, no. 2, pp. 158–166, 2016.

[11] Y. Ma, Z. Li, and M. A. Sotelo, “Testing and evaluating driverless vehicles’s intelligence: The Tsinghua lion case study,” *Intell. Transp. Syst. Mag.*, vol. 12, no. 4, pp. 10–22, 2020.

[12] T. L. Saaty, *The Analytical Hierarchy Process: Planning, Priority Setting, Resource Allocation*. Pittsburgh, USA: RWS Publish, 1990.

[13] SAE, “J3016 levels of driving automation,” *SAE Official Website*, pp. https://www.sae.org/news/2019/01/sae–updates–j3016–automated–driving–graphic, 2019.

[14] F. E. Schneider, D. Wildermuth, and W. H. L., “Professional ground robotic competitions from an educational perspective: A consideration using the example of the European Land robot trial (ELROB),” in *Proc. of 6th IEEE Inter. Conf. Intel. Syst.*, Sofia, Bulgaria, 2012, pp. 399–405.

[15] L. Staplin, T. Mastromatto, K. H. Lococo, W. Kenneth, K. W. Gish, and J. O. Brooks, “The effects of medical conditions on driving performance,” in *Peprot No. DOT HS 812 623*, Washington, DC: National Highway Traffic Safety Administration, 2018.

[16] Y. Sun, G. Tao, G. Xiong, and H. Chen, “The fuzzy-AHP evaluation method for unmanned ground vehicles,” *Appl. Math. Inf. Sci.*, vol. 7, no. 2, pp. 653–658, 2013.

[17] Y. Sun, G. Xiong, H. Chen, and S. Wu, “A cost function-oriented quantitative evaluation method for unmanned ground vehicles,” *Adv. Mat. Res.*, vol. 301-303, pp. 701–706, 2011.

[18] Y. Sun, G. Xiong, X. Ma, J. Gong, and H. Chen, “Quantitative evaluation of unmanned ground vehicle trajectory based on chao theory,” *Adv. Mech. Eng.*, vol. 7, no. 11, pp. 1–9, 2015.

[19] Y. Wang, K. Li, Y. Hu, and H. Chen, “Modeling and quantitative assessment of environment complexity for autonomous vehicles,” in *Proc. of the 2020 CCDC*, Hefei, China, 2020, pp. 2124–2129.

[20] Y. Zhao, K. Meng, and L. Gao, “The entropy-cost function evaluation method for unmanned ground vehicles,” *Math. Probl. Eng.*, pp. 1–6, 2015, ID 410796.