Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

Human-AI Collaboration via Hybrid Intelligent System for Safe Autonomous Driving

J. Jayapradha[1], Dr. Ben Sujin. B[2], Dr.M.Janaki Rani[3], Rayappan Lotus[4], A Fayaz Ahamed[5]

[1]*Department of Computing Technologies, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu 603203, India. jayapraj@srmist.edu.in*
[2]*PhD, Faculty and Researcher, Computer Engineering, University of Technology and Applied Sciences – Nizwa, Sultanate of Oman. bensujin.bennet@utas.edu.om*
[3]*Professor, ECE Department, Dr.M.G.R Educational and Research Institute, Maduravoyal, Chennai 95. janakirani.ece@drmgrdu.ac.in*
[4]*Assistant Professor, Computer Science and Engineering, Vel Tech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology, Avadi, Chennai. sjlotusanton@gmail.com*
[5]*Assistant Professor, R.M.K Engineering College, Kavaraipettai, Tamilnadu, India. afd.eee@rmkec.ac.in*

The growing penetration of artificial intelligence in applications including autonomous vehicles and brilliant factory equipment amplifies the risk of catastrophic damages from the execution of incomplete machine learning algorithms. For providing durable and dependable effectiveness, Artificial Intelligence (AI) frameworks require massive quantities of thoughtfully curated training data. However, because the production of training data frequently wants trained manual annotation, which limits scalability, it is scarce. Thus in this research, we propose a hybrid intelligent system for a human-machine collaborative environment. This can help humans interpret the machine and finish any assignments on budget. Considering not all automated machines can manage occupations by themselves, this human-AI shared effort will have an immense effect on many different sectors. Human-machine intelligence is blended into human-in-the-loop computing to establish a hybrid intelligence regarding supplementary habits. Humans play a role with their dynamic and creative mental abilities, but algorithms are unparalleled in logic and calculation speed. To acquire high accuracy and confidence in machine learning methodologies, hybrid cognitive systems are mandatory. To make certain that potential applications are successful and dependable, designers must establish and reinforce the trust between humans and AI, starting with being an AI-ready organization and ending with a clear focus on the benefits a hybrid intelligent system can offer customers. The alliance connecting humans and technology will become even larger because of the developing trends of technical innovations and the introduction of artificial intelligence in a growing range of applications.

Keywords: Human-AI collaboration; Artificial intelligence; hybrid intelligent system; safe

driving of autonomous vehicle.

1. Introduction

One of the most intriguing approaches to advocating smart mobility is autonomous driving, which drastically reduces the potential risks correlated with human behavior and driver drowsiness. Sensors mounted in the conveyance to keep an eye on nearby objects are vital boosters for autonomous driving systems. Additionally, prediction and forecast models leverage sensor data to discover the existing driving state and find out the beneficial route of action. To lessen the likelihood of compromising the lives of road actors, the models require to be quite precise and have a minimal processing time. In this application arena, supervised learning performs far superior to routine identification algorithms. However, a supervised identification model needs immense constitutes of training data to ultimately identify objects in a potent, accurate, and verified way. A high-accuracy model for figuring out objects in the street requirements to comprehending and take consideration numerous factors, encompassing rain, sunlight, sunsets, midnight, and seasons, and all four of them require particular parameters for elements like illumination and reflectivity. Meanwhile, public, precise, reliable, and particularly large quantities of data are challenging to obtain for particular targets, and the handful that are offered aren't useful for certain level purposes, making it problematic to derive conclusions from these people. In light of these severe limitations and failures in high-scale situations picking up objects for autonomous driving is still an unresolved challenge. In this research, we explain our initial efforts toward a Human-AI union to help rapid but incredibly precise camera portray segmentation for autonomous driving. We reveal domain-specific obstacles with image information and the associated labels, and we discuss how to properly take these issues into thought while composing the crowdsourcing strategy. To decrease the crowdworkers' occupation and assist them when they deal with a lot of data, an AI model produces pre-annotations of the shots. In a user study, crowdworkers who labeled more than 500 real-world photos designed the strategy [1].

The collection of hardware and software innovations designated the human-machine interface, or HMI, promotes communication between autonomous vehicles (AVs) as well as humans accessing the road, consisting of passengers. Effective communication and trust can be fostered in this dynamic tandem by well-designed HMIs that broadcast information, commands, and objectives both inside and outside of the vehicle. Investigating the human factor elements of AV interfaces is essential when full self-driving SAE Level 5 continues to be developed. Figure 1.1 gives the simple architecture of Hybrid human-artificial intelligence [2].

The first dimension that demands to be declared for the reason to put up hybrid intelligence systems is the task that requires a solution. Four broad-form task categories have been defined in this context: action, reasoning, prediction, and recognition. In the beginning, procedures that belong, assume, objects, images, or natural language can be defined as recognition tasks. These particular kinds of maneuvers are utilized in self-driving automobiles as well as alongside automated assistants like Duplex, Siri, or Alexa. Afterward, prediction tasks are supposed to plan potential events by using historical information such as market dynamics or stock prices. Reasoning, the third task category, focuses on comprehending facts by, for example, inductively establishing (mental) models of identifiable issues. It enables one to grapple with tricky issues with a tiny quantity of data [3].

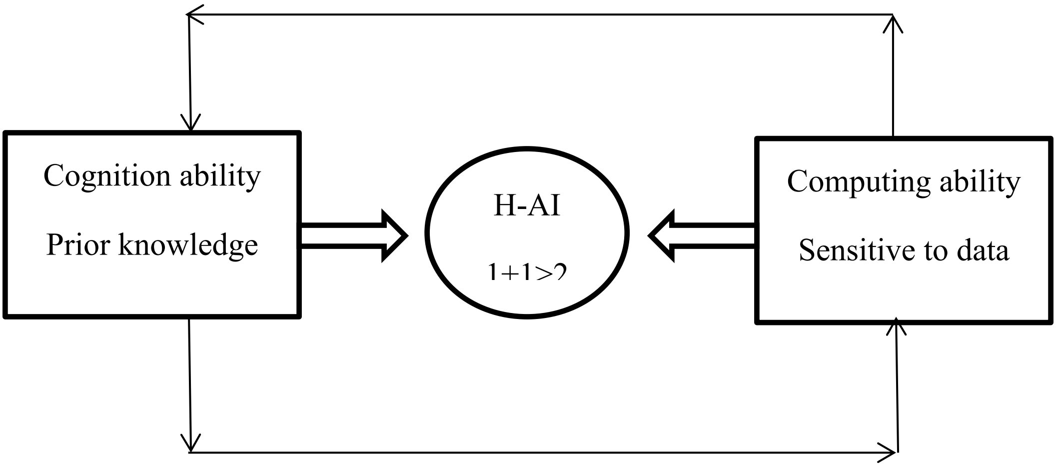

Fig. 1.1. Hybrid Human-Artificial Intelligence

Difficulties with large-scale execution encompass steep expenses, moral and legal difficulties, and technological issues. Thus, asking for the utmost security mandates a blend of AI and the cognitive abilities of humans. It seems as though humans have potential benefits incorporating adaptation, flexibility, and beneficial adjustment, but when driving, they are more inclined to become influenced by sensations, drowsiness, or inclement weather. Studies on partnered driving in real-life situations are beginning to focus on the associated positive effects and pitfalls. For instance, many magazines highlighted Bluetooth's data and information fusion gadgets, which endeavor to fully understand the natural world when driving. These items supplied an integrated driving semi-physical simulation system that enabled machines and drivers to control the cars to ensure public safety. We envision that the seamless fusion of AI and human intelligence in H-AI will drastically boost the intelligence of transportation. With the use of various electronic devices, drivers would be able to sense situations more accurately and make better decisions. Along with that, machines would be able to replace human drivers in potentially hazardous situations by originating synchronized intuitions and cognitive abilities from human drivers [4].

The chronological arrangement of the article is as follows: An extensive evaluation of the hybrid H-AI for safe autonomous driving will be provided in Section 3. The experimental results will be examined in Section 4; the navigation, mapping, and scene processing capacities of the robot will be described in detail and highlighted in Section 5; subsequently, a technologically sophisticated robotic system for indoor surveillance will be offered in Section 5.

2. Related Works

[5] This position paper attempts to grant the AI and automated driving social groups an extensive overview of the barriers and possible futures for research about the issue of intriguing assignment among the driver and the vehicle in the environment of autonomous automobiles. The vehicle's Command System is utilized to ensure that sure that information about the driver and the vehicle is truthful that multiple sources of data are efficient, and that it appears context and situational awareness are readily accessible to navigate the unanticipated events that could come up in advance of and throughout the dynamic delegation phase. It requires attention to craft and finding adaptive and non-intrusive Human-Computer Interaction procedures to allow the driver an adapted and lightning-fast control transition to the vehicle. These interaction methods that requires be smooth and shape-changing, and they must be reinforced by numerous kinds of homemade, multimodal interaction mechanisms, encompassing speech, gestures, gaze, and haptics. The driver will also be reminded in instances of an unforeseen or harmful instance with the help of these multimodal electronics.

[6] We researched an increasing pool of scientific literature published in peer-reviewed works integrating risk and safety science, human components and training, marine policy, and ocean and reliability engineering to outline the state-of-the-art research for autonomous ship systems in this article. Understanding that shipping has become more sophisticated, we put forward three coursework questions: (1) How have autonomous ship systems currently enforce human control? (2) What theories, mechanisms, and treatments are being adopted to overcome design flaws and safety concerns? and (3) which deficiencies in research, regulatory obstacles, and glitches hold in the method of their actual seen procedure. We gathered information generated by 42 pertinent, reviewed papers in a systematic review. Two fundamental scenarios—safety and control—underpinned every verdict despite the review's broad coverage throughout five distinct fields (marine policy, ocean engineering, psychology, reliability engineering, and risk science). The two subject matters were closely linked, with safety being characterized as a property that arises from ensuring machine and human autonomy in the context of control. In response to this point of view, human autonomy looks after supervisory goals in the presence of unclear and fluctuating outside variables, in comparison to machine autonomy implements previously established chores.

[7] A roadblock to acceptance is getting larger as further ASVs are installed and AI-based navigational gadgets emerge. This is due to the fact citizens are growing cautious surrounding trust. This is exclusively acute as we possess curbside ASVs, where prototypical shortcomings could result in unforeseeable possibilities that endanger people. With a primary emphasis on feasibility for interpretation, understandability, explainability, and trust, the fundamental values of Explainable AI (XAI) represent a common ideal for various fields interested in dealing with the issues of fostering enthusiasm in ASVs. Computer scientists typically set up XAI, with a view of promoting model accuracy and efficiency using a superior construal of opaque, "black box" machine learning (ML) models. There are several kinds of important regulations. First and foremost, the mechanisms that we suggest as "human-centered XAI" are susceptible to transformation over time, through frontiers of culture, and in association with progressing cultural norms, as concepts as a result of their very essence are adaptable. Second, each example that had been presented was taken from

our investigations or the research that our university's collaborators carried out. This necessitated an extensive degree of consciousness in this research process, which faced the risk of frustrating the processes when we realized how our concept had been created. This work has achieved that design methodologies are changing towards end-user interaction design as ASVs scale for mass adoption. But not a single of this is taking place based on the umbrella of an adequately developed sense of sight of what the process of design takes up. This work's first-degree contribution is the formulation and definition of human-centered XAI, which progresses ASV design toward mainstream adoption.

[8] The advent of AI has altered our lives and our expectations of the future. Machines containing intelligence have become human companions. People are surfing for new computational models and approaches to utilize AI in the recognition of ubiquitous computing and smart machines. Hybrid-augmented intelligence is one of the primary directions for the future development of AI. Building human-computer connection-based HITL hybrid-augmented knowledge by combining perception and cognitive aptitudes of mankind with the device's potential to figure out and retrieve information can enormously enhance the AI system's choice proficiency, the different levels of memory sophistication necessitated to handle multifaceted tasks and responsiveness to complex situations. Employing experience learning, intuitive reasoning, and other hybrid models, hybrid-augmented cognitive ability based on CC might tackle the long-lasting planning and reasoning challenges in the AI study domain.

[9] We outlined the Driver AdvocateTM (DA) system's architecture, underscoring the manner machine learning and agent technologies are linked together. A high-fidelity driving simulator functions as the starting point for a prototype system that incorporates some parts of the architecture. The prototype DA was previously employed in human driving investigations to ensure that it can be further programmed to modify itself to individual dissimilarities in the driving approach. Whenever the DA reveals the mandatory equipment, we are going to test the system in a real car under the real driving scenarios. The tremendous as well as varied broad spectrum data and knowledge that the DA requires handling puts forward another obstacle. This stretches from numerical or statistical data (which includes driving logs accumulated from the simulator, control data being sent to the vehicle's actuators, etc.) to symbolic knowledge (which is necessary to describe task models and for planning scenarios and assessments). Although each agent can multiply together with its representation of the regulation courtesy to our architecture, several agents need to collaborate on some kind of hybrid reasoning their potential. Again, the DA's enthusiasm for real-time implies that the efficiency issue could turn out to be vital.

[10] In this paper, we introduce an uncommon consumption mechanism of human pointing and produce a reinforcement learning framework based on human guidance. We next propose a technique, named PHIL-TD3, without the objective of strengthening algorithmic competencies underneath the context of human-in-the-loop reinforcement teaching. To ease the financial strain on individuals, we additionally present a mechanism for emulating human behavior. Two demanding autonomous driving tasks were assigned to PHIL-TD3, and the outcomes of its work are juxtaposed with the identical non-guidance baseline and the highly powerful human-guidance-based RLs. The experimental data generate three primary outcomes: 1) The hypothesized PHIL-TD3 could surpass state-of-the-art human-guidance-

based RLs about asymptotic performance, enhancing efficacy in learning by over 700% and 120% in the two eventualities that are deployed, respectively. 2) In a multitude of requirements, the chosen PHIL-TD3 exceeds the other model devices in terms of performance, toughness, and versatility when compared to the two demanding autonomous driving tasks. 3) The postulated T DQA procedure delivers an important improvement to the advancement of PHIL-TD3 and has the potential of accurately assessing the variety of various surveillance by humans, eliminating humans of a portion of their burden.

[11] Perception, prediction, and control tasks were especially tricky in AD systems. DRL is an attainable likelihood for the future development of AD systems, potentially allowing was required behaviours to be explained first in emulation and additional refined on real datasets, instead of being explicitly programmed. In this publish, we provide an introduction of AD system sections, DRL methodologies and implementations of DRL for AD. We discuss over the main challenges that ought to be overcome for the reason to enable the DRL to be increasingly and realistically implemented in AD applications. In contrast a great deal of the study's subjects discussed in this research report has been performed in simulated locations the arrival of applications on practical motor vehicle is interesting. The primary obstacles for creating an effective real-world system would be surpassing crucial challenges include safety in reinforcement therapy, upgrading data efficiency, and ultimately allowing transfer learning along simulated conditions.

[12] The current study introduces an unusual reinforcement learning algorithm utilizing expert demonstrations for the consumption benefit of human prior expertise to enhance the performance and efficiency of the experiments. More specifically, we pair up copying the expert's endeavors and improving the Q-function to update the policy network. We additionally formulate an adaptive experience replay handle which permits our team to adaptively sample experience from the agent's self-exploration and specialist demonstration for revising the policy. We investigate the recommended method in a costly, traffic-heavy, simulated urban roundabout. A careful examination of different RL and IL baselines demonstrates that our technique functions better in the training process with regard to sample efficiency. The assessment findings suggest that the suggested strategy helps get to the destination quickly and with a higher degree of success amount. We further demonstrated that the success rate is capable of being boosted even more by pairing the RL-based controller with a rule-based safety controller.

3. Methods and Materials

In conventional DRL applications, for instance autonomous driving, the control of the DRL agent can be characterized as a Markov decision process (MDP). The MDP is represented by a tuple N which includes the state space $T \in \mathbb{S}^o$, action space $B \in \mathbb{S}^n$, transition model $U: T \times B \rightarrow T$, reward function $S: T \times B \rightarrow \mathbb{S}$, and state space (where o and n are the multifaceted of the state space and action room, respectively; $\mathbb{S}$ is the real number set).

$$N = (T, B, U, S) \qquad (1)$$

The agent undertakes an action $b_u \in B$ in a state $t_u \in T$ at a time step u, and it generates a reward signal$s_u = S(t_u, b_u)$. Then, in compliance with the environmental dynamics

$U(\cdot\,|t_u, b_u)$; at, the environment develops into a next-step state $t_{u+1} \in T$. It is complicated to formulate the transition hazard model U for the surrounding dynamics in the autonomous driving scenario. As a conclusion, we utilized model-free reinforcement instruction to manage this objection, which avoids the necessity of grasp the transition dynamics.

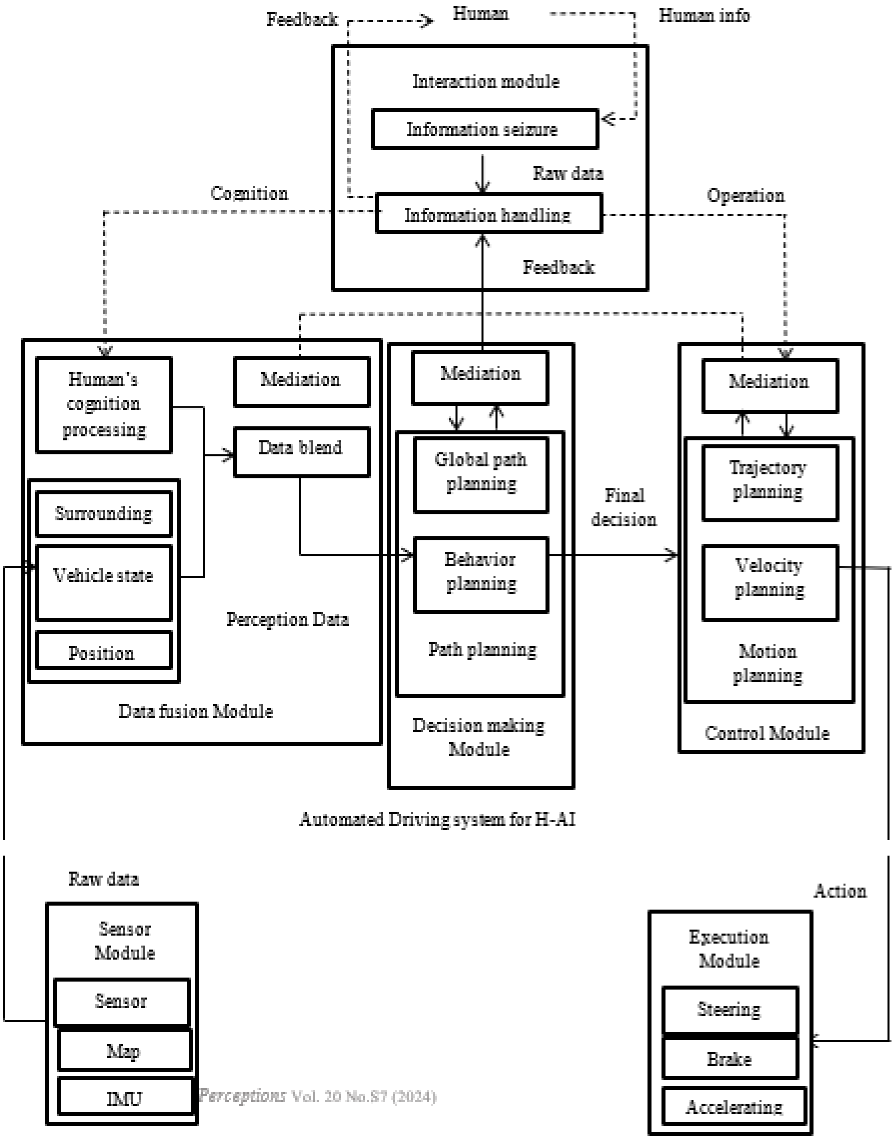

Fig. 3.1. Demonstrates the general development of the H-AI autonomous driving system.

The most significant of the primary difficulties for semi-autonomous driving is always HMC. We suggest that conflicts happen because of inappropriate human-machine coupling. As an example, Figure 3.1 demonstrates the theoretical architecture whose services are based on H-AI. General autonomous driving systems have the luxury of understanding, selecting, and regulating. Consider that we separately segregate human-machine interaction as an autonomous part due to the fact it is extremely crucial to the broader system. The high-level architecture in the present investigation is constructed with an exceptionally comprehensive off-policy actor-citric method, TD3. Under the pressure of value network R, the TD3 algorithm updates its action-selection policy as well as determines an unpredictable action using policy network ρ. Depending on the Bellman iterative equation, the value matrix approximates the value of the individual state and action. Next, with the goal to cope with the overestimation issue along with TD3 forms two value networks, R1 and R2. The target networks ρ', R_1', and R_2' are implemented to promote learning. For the development of the human-in-the-loop building structure that lie within our reinforcement learning algorithm, we combine LfD and LfI into an equivalent creating it. With that arrangement, individuals possess the flexibility to determine when they'll enter in, overturn the first rule selection, and display scenarios in real time. The end result develops an algorithm for online transition spanning agent discovery and human oversight.

Let $I(T_u) \in \mathbb{S}^o$ symbolize the individual's policy. The guidance for human intervention is generated employing a random event procedure $J(T_u)$, considering into thought the driver's observation of the current parameters. Agent action b_u should therefore be represented like this:

$$b_u = J(T_u) \cdot b_u^{human} + [1 - J(T_u)] \cdot b_u^{DRL} \quad (2)$$

$$b_u^{DRL} = clip\big(\rho(T_u|\Xi^\rho) + clip(\theta, -D, D), b_{low}, b_{high}\big), \in \sim O(0, \tau) \quad (3)$$

where $b_u^{human} \in I$ is the human-given guidance action, b_u^{DRL} is the policy network's action, $J(T_u)$ is equal to 0 in the absence of human guidance or 1 in the presence of human action, indicates Ξ^ρ the policy network's parameters, b_{low} and b_{high} are the action space's lower and upper bounds, respectively, θ is the noise subject to a Gaussian distribution with a τ standard deviation, and d is the clipped noise boundary. To encourage research in the deterministic policy, Gaussian noise is stipulated. When a human participant chooses to intervene in an occurrence during agent training, the mechanism established by Equation (2) fully transfers the driving control authority to them. The value function, which is calculated from the assumption of future reward in a similar way, is roughly given by the value network:

$$R^\pi(t, b) = \mathbb{F}_{t \sim U, b \sim \pi(\cdot|T)}\left[\sum_{j=0}^{\infty} \delta^j . s_j\right] \quad (3)$$

where $\mathbb{F}[\cdot]\sim$ represents the mathematical expectation, j is the index of the counted time step, and δ is the reduction factor used to determine the relevance of future rewards. Assuming that the simplified variant of $R^\pi(t, b)$ is $R(t, b)$. Unless specified differently, the superscript regarding the policy π is neglected.

The Bellman iteration is applied to solve the above expectation, and the predicted repetitive target of value function z at step u can be determined as follows:

$$z_u = s_u + \delta \min_{k=1,2} R_k'\left(T_{u+1}, \rho'\left(T_{u+1}|\Xi^{\rho'}\right)|\Xi^{R_k'}\right) \quad (4)$$

Where k projects the index of two value networks, R_1 and R_2 , and $\Xi^{\rho'}$ denotes the variables of the target policy network and, $\Xi^{R_k'}$ the specifications of the target value networks.

$$M^{R_k}(\Xi^{R_k}) = \mathbb{F}_{(T_u,b_u,s_u,t_{u+1})\sim E}[\|z_u - R_k(T_u,b_u|\Xi^{R_k})\|^2] \quad (5)$$

where $\mathbb{F}$ corresponds to the experience replay buffer E, which is comprised of the present condition of the action, the reward, and the state of the next step; Ξ^{R_k} corresponds to the expected value; and Ξ^{R_k} is standing for the parameters of the value networks.

To effectively utilize the economic value of the value network, or to enhance control efficiency in the designated autonomous driving scenario for this study, is the primary objective of the policy network that decides the control action. As the outcome, the policy network's loss function in the TD3 algorithm is developed as follows:

$$M^{\rho}(\Xi^{\rho}) = -\mathbb{F}[R_1(T_u,b_u^{DRL})] = -\mathbb{F}_{T_u\sim E}[T_u,\rho(T_u|\Xi^{\rho})] \quad (6)$$

About Equation (6), the primary objective of the policy is to mitigate the policy network's loss function while enhancing the value of the value network. Since the noise in Equation (2) displays a zero-mean distribution, the unbiased assessment of b_u^{DRL} is equivalent to that of $\rho(T_u|\Xi^{\rho})$.

The TD3 algorithm's loss function requires to be altered to consider the human experience when human guidance (b_u^{human}) appears. Therefore, the result for the network in Equation (5) can be formulated in the structure that follows:

$$M^{R_k}(\Xi^{R_k}) = \mathbb{F}_{(T_u,b_u,s_u,t_{u+1})\sim E}\left[\left(z_u - R_k\left(T_u,b_u^{human}|\Xi^{R_k}\right)\right)^2\right] \quad (7)$$

For the reason of trying to get rid of the previously referred to inconsistency issue, we add a human guidance term J to the policy network's loss function that's shown in Equation (6).

$$M^{\rho}(\Xi^{\rho}) = \mathbb{F}_{(T_u,b_u,J(T_u))\sim E}\{-R_1(T_u,b_u) + J(T_u).\varphi_J.[b_u - \rho(T_u| = \Xi^{\rho})]^2\} \quad (8)$$

where φ_J is a factor that adjusts the human supervision loss's weight; b_u^{DRL} in Equation (6) may consequently be exchanged with at to compensate for both DRL policy actions and human actions. When human advising happens, the changed direction is in alignment with $\{T_u, b_u^{human}\}$.

Thus, it is suitable to generate a tolerant task mechanism for factor φ_J that is associated with the trustworthiness of human endeavors. To do this, we recommend the R-advantage as a permitted evaluation metric, and the postulated weighting factor can be changed as follows:

$$\varphi_J = \omega^l \cdot \left\{max\left[exp\left(R_1(T_u,b_u) - R_1\left(T_u,\rho(T_u|\Xi^{\rho})\right)\right),1\right] - 1\right\} \quad (9)$$

Where l is the learning episode index and ω is an extreme parameter that is slightly smaller than 1. The chronological decay factor ω^l illustrates that as the policy function continuously achieves adulthood, human guidance turns less precise.

The policy network's batch gradient can be measured through Equation (9).

$$\Delta_{\Xi^{\rho}} M(\Xi^{\rho}) = \frac{1}{O}\sum_{u=1}^{O}\{\left(-\Delta_a R_1(T,b)|_{T=T_u,b=\omega(T_u)}\Delta_{\Xi^{\rho}}\rho(T)|_{T=T_u}\right) + \left(\Delta_{\Xi^{\rho}}\left(\varphi_J \cdot \|b - \rho(T)\|^2\right)|_{T=T_u,b=b_u}\right) \cdot J(T_u)\} \quad (10)$$

where N corresponds to the experience replay buffer E's batch size sample.

Ultimately the human advising component is introduced as follows in which the experience replay buffer's the beginning stored tuple is tweaked:

$$E = \{T_u, b_u, s_u, T_{u+1}, J(T_u)\} \quad (11)$$

The outcome is in a reworked DRL algorithm that features human guidance attainable in real-time [13].

4. Implementation and Results

4.1 An overview of the attempts

A human-in-the-loop driving simulator was employed for a series of experiments dealing with 40 people participating in the envisioned autonomous driving scenarios to evaluate the viability and practicality of the hypothesized improved DRL with human direction. There were six typical scenarios for everybody; one was for the proposed method's training process (associated with Experiment A in the form of E), and the additional five were created to assess and criticize the designed algorithm's performance. The training scenario incorporated an intricate driving movement such as unaltered lane shifting and overtaking, as they were the environment's adapted reward emphasized cautious and transportable driving. The ego vehicle is required to emerge from the spawn location, stay underneath the road, avoid hitting any other obstacles, and then ultimately reach the finishing line to successfully perform the tasks as envisioned in all circumstances. The current episode concluded instantaneously if the ego vehicle smashed with the boundary of the road or other vehicles in traffic. A new episode might start with additional generating vehicles concentrating on the training process. In the testing circumstances, the types, jobs, and speeds of the surrounding objects are modified to boost the policies' training performance under various conditions with greater demands. The ego vehicle required to emerge from the spawn location, stay underneath the road, avoid hitting any other obstacles, and then ultimately reach Experiment A was conducted to validate the training performance improvement by considering the proposed approach with alternate human-guidance-based DRL methodologies. For the sake of comparison, we initially developed all associated baseline DRL algorithms employing the same form of real-time human guidance. To be clearer, there are three baseline approaches to DRL: the Vanilla-DRL procedure (the standard TD3 algorithm without human guidance), the HIRL (shaped value function, but no modification to the policy function), and the IARL (fixed weighting factor φ_J pertaining to human guidance in the policy function of DRL). To promote speedier convergence, supervised learning pre-initializes every single policy network in these. Please pay attention to the Method Section for wholehearted implementations of the stated concepts.

4.2 The boosted instruction efficacy of the postulated Hug-DRL strategy

The outcomes illustrated in Figure 4.1, drawn from Experiment A, correspond to the increase in performance that the proposed Hug-DRL method brought when contrasted to other advanced human-guidance-based algorithms, among them the Intervention-Aided DRL (IARL) and Human-intervention DRL (HI-RL), and in addition to the Vanilla DRL without human guidance (a pure TD3 algorithm). In the context of the experiments, the timestep reward obtained and the length of time of each episode were recorded and assessed for each participant, to assess the training performance throughout a training session under each method. Both the episodic reward and the duration of the episode were assessed, as shown in Figure 4.1 and 4.2. The observations suggest the proposed Hug-DRL method exceeded all other baseline methods. Table 1 shows the training of the different DRL methods corresponding to each step.

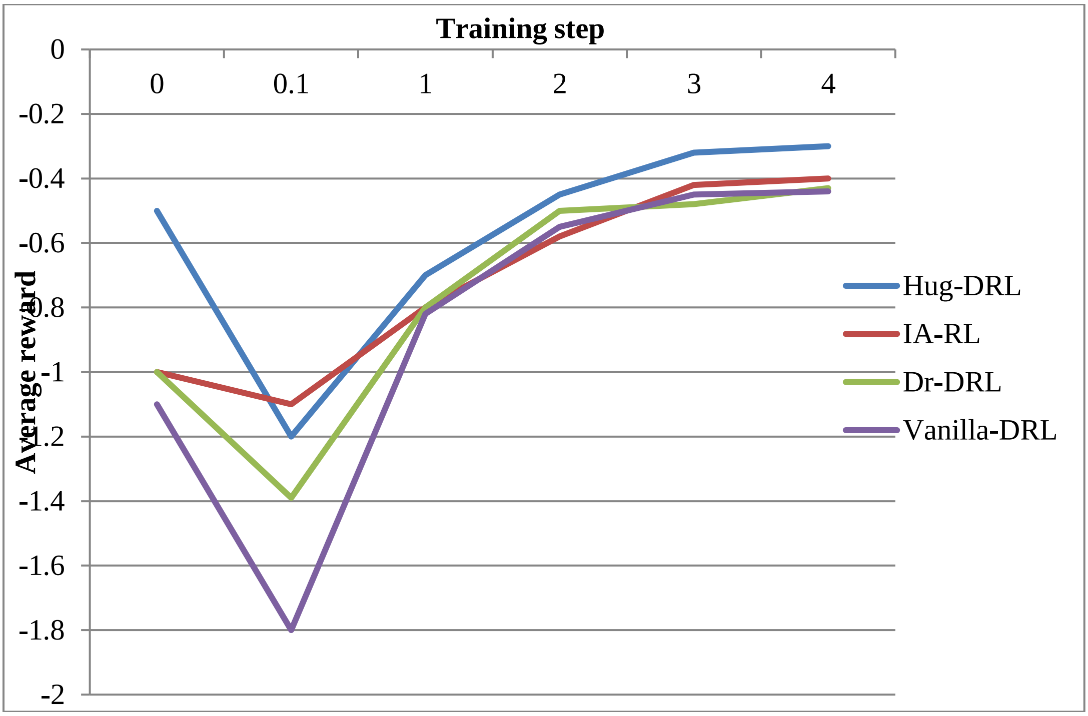

Fig. 4.1. Consequences of the different techniques utilised for the episodic training reward.

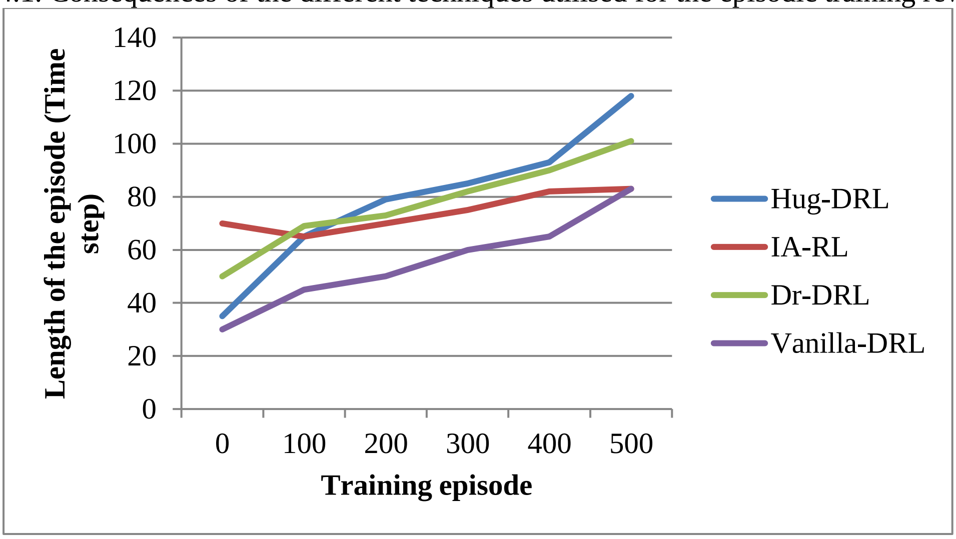

Fig. 4.2. Consequences of the Three Techniques for the Episodic Length.

Table 1. Illustration of Training According to the Various DRL Methods

Training steps	Four different methods of DRL			
	Hug-DRL	IA-RL	HI-RL	Vanilla DRL
0	-0.5	-1	-1	-1.1
0.1	-1.2	-1.1	-1.39	-1.8
1	-0.7	-0.8	-0.8	-0.82
2	-0.45	-0.58	-0.5	-0.55
3	-0.32	-0.42	-0.48	-0.45
4	-0.3	-0.4	-0.43	-0.44

Table 2. Representation of number of training episodes for the particular AI approach

Training episode	Four different methods of AI			
	Hug-DRL	IA-RL	HI-RL	Vanilla DRL
0	35	70	50	30
100	65	65	69	45
200	79	70	73	50
300	85	75	82	60
400	93	82	90	65
500	118	83	101	83

The averaged reward obtained with the proposed approach right through the whole training procedure was the highest at (N = -0.599, TE = 0.042), by an examination of the statistical results shown in Fig. 3c. This was followed by the monetary rewards obtained with the IA-RL (M = -0.660, SD = 0.033), the HI-RL method (M = -0.697, SD = 0.029), and the Vanilla-DRL method (M = -0.688, SD = 0.052). Additionally, the one-way ANOVA analysis provided by Supplementary Table 2 suggested that these variations were statistically significant, and $F(5,37)=28.98$. Furthermore, the entire length of the episode—which specifically determines the ability to achieve the task—was also analyzed in between all three strategies. The recommended method's mean value (N = 86.2, TE = 6.3) was more effective when compared to the HI-RL method's (N = 82.6, TE = 7.3), the IARL method's (N = 89.0, TE = 9.5), and the Vanilla-DRL method's (N = 65.2, SD = 8.6) based on the analysis of the data presented in Fig. 3d. $F(5,37) =15.05$, and the statistical significance of their differences was also established, as demonstrated by the ANOVA analysis. Considering the asymptotic rewards, the recommendations for Hug-DRL, IA-RL, and HI-RL provide improved performance of 31.9%, 14.2%, and 7.1%, respectively, as opposed to Vanilla-DRL. The previous results highlight the extent to which human involvement might enhance DRL performance [14].

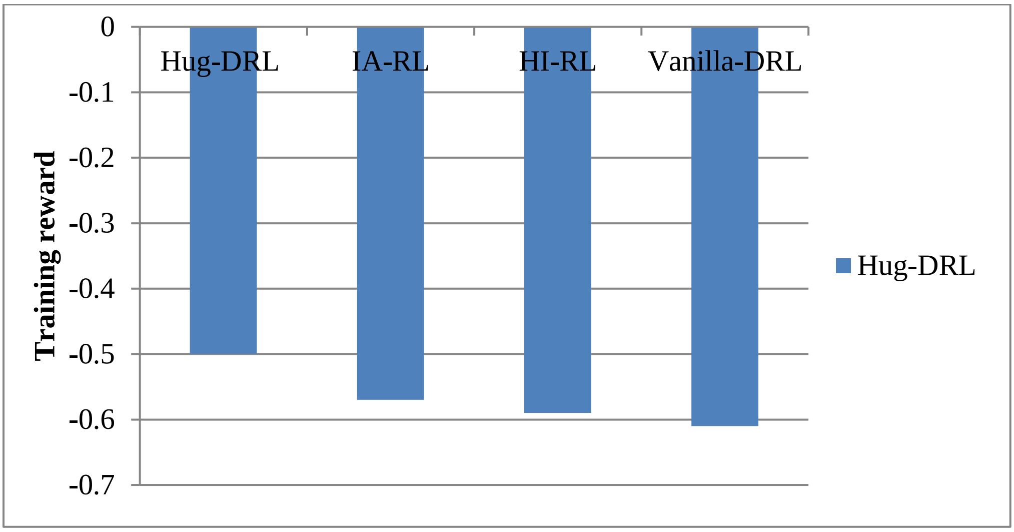

Fig. 4.3. Observations from the average reward all across the period of a training session according to different approaches.

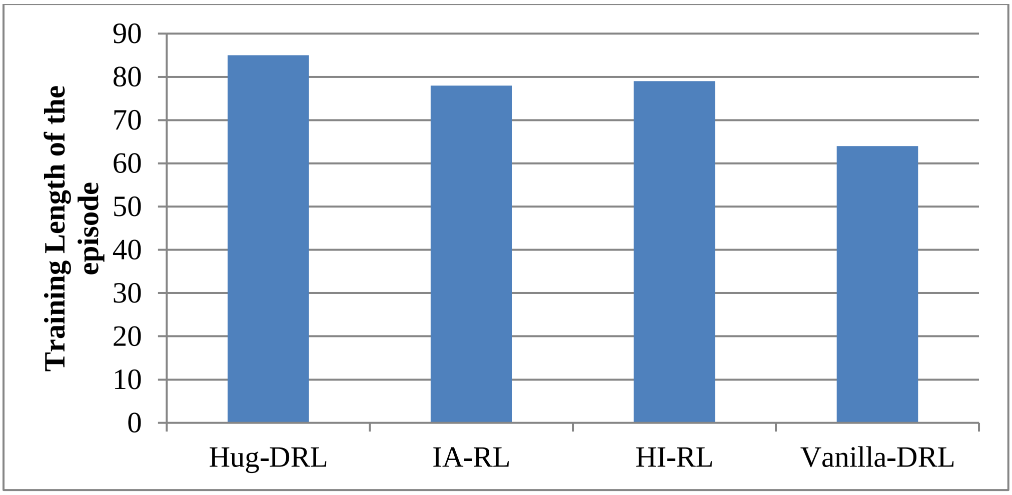

Fig. 4.4. Outcomes on an average period among occurrences during the span of the training session by implementing various strategies

5. Conclusion

A real-time Hug-DRL methodology was established in this study to train principles in an end-to-end self-driving technology instance. A modified version of the actor-critic architecture encompassing revolutionized policy and value networks was drawn up. During the training phase, humans could come in and correct the agent's erroneous DRL actions in real-time. The new approach was compared with other state-of-the-art educational frameworks and validated with forty test respondents through human-in-the-loop tests.

According to the experimental data, it's evident that the recommended Hug-DRL carries out significantly on examinations than present-day arrives concerning learning efficiency. Both the initial education and online fine-tuning periods of the agent's training performance can be significantly improved by the approach recommended. Given that intermittent human involvement has little impact on the load for mankind, it can be an effective means of boosting DRL performance. More precisely, the suggested method greatly reduces the challenges on the human side. Individuals don't have to be consultants with more thorough expertise or understanding of particular disciplines. Even though human acts are undesirable, the DRL can be securely learned and widened as long as it cooperates normally and applies common sense. The recommended approach exhibits immense potential for implementation in emerging real-life applications due to these reasons. The high-level framework, the strategies implemented, and the algorithms developed in this work present quite a bit of potential for advancement into additional fields incorporating AI and human-AI interaction.

References

1. Gamboa, E., Libreros, A., Hirth, M., & Dubiner, D. (2022, October). Human-AI Collaboration for Improving the Identification of Cars for Autonomous Driving. In *CIKM Workshops*.
2. Meiselwitz, G. (2015). *Social Computing and Social Media*. Springer International Publishing.
3. Dellermann, D., Calma, A., Lipusch, N., Weber, T., Weigel, S., & Ebel, P. (2021). The future of human-AI collaboration: a taxonomy of design knowledge for hybrid intelligence systems. *arXiv preprint arXiv:2105.03354*.
4. Shi, F., Zhou, F., Liu, H., Chen, L., & Ning, H. (2022). Survey and Tutorial on Hybrid Human-Artificial Intelligence. *Tsinghua Science and Technology*, *28*(3), 486-499.
5. Lemercier, C., Combettes, S., Jauregui, D. G., Lagasquié-Schiex, M. C., Dubois, E., & Couture, N. (2019, September). Adaptive Human Machine Interfaces in an Autonomous Vehicle. In *Workshop on Explainable AI in Automated Driving: a User-Centered Interaction Approach in 11th International ACM Conference on Automotive User Interfaces and Interactive Vehicular Applications (Automotive UI 2019)* (pp. 1-9).
6. Veitch, E., & Alsos, O. A. (2022). A systematic review of human-AI interaction in autonomous ship systems. *Safety science*, *152*, 105778.
7. Veitch, E., & Alsos, O. A. (2021). Human-centered explainable artificial intelligence for marine autonomous surface vehicles. *Journal of Marine Science and Engineering*, *9*(11), 1227.
8. Zheng, N. N., Liu, Z. Y., Ren, P. J., Ma, Y. Q., Chen, S. T., Yu, S. Y., ... & Wang, F. Y. (2017). Hybrid-augmented intelligence: collaboration and cognition. *Frontiers of Information Technology & Electronic Engineering*, *18*(2), 153-179.
9. Hwang, C. H., Massey, N., Miller, B. W., & Torkkola, K. (2003). Hybrid Intelligence for Driver Assistance. In *FLAIRS* (pp. 281-285).
10. Wu, J., Huang, Z., Huang, W., & Lv, C. (2022). Prioritized experience-based reinforcement learning with human guidance for autonomous driving. *IEEE Transactions on Neural Networks and Learning Systems*.
11. Talpaert, V., Sobh, I., Kiran, B. R., Mannion, P., Yogamani, S., El-Sallab, A., & Perez, P. (2019). Exploring applications of deep reinforcement learning for real-world autonomous driving systems. *arXiv preprint arXiv:1901.01536*.
12. Liu, H., Huang, Z., Wu, J., & Lv, C. (2022, June). Improved deep reinforcement learning with expert demonstrations for urban autonomous driving. In *2022 IEEE Intelligent Vehicles Symposium (IV)* (pp. 921-928). IEEE.

13. Wu, J., Huang, Z., Hu, Z., & Lv, C. (2023). Toward human-in-the-loop AI: Enhancing deep reinforcement learning via real-time human guidance for autonomous driving. *Engineering*, *21*, 75-91.
14. Wu, J., Huang, Z., Huang, C., Hu, Z., Hang, P., Xing, Y., & Lv, C. Human-in-the-loop deep reinforcement learning with application to autonomous driving. arXiv 2021. *arXiv preprint arXiv:2104.07246*.