Autonomous Vehicles Through the Prism of the Turing Test

Anna A. Gorbenko, Vladimir Yu. Popov

***Abstract*— In recent years, significant progress has been made in research on various problems of autonomous transport. According to most forecasts, autonomous vehicles will appear on the roads in the coming years. However, strict requirements for the artificial intelligence of autonomous vehicles have not yet been formulated. We do not yet have a clear understanding of the intelligence for autonomous vehicles. Nevertheless, the problem of developing an analogue of the Turing test for autonomous vehicles has attracted increasing attention in recent years. There are a number of different points of view on the analogue of the Turing test for autonomous vehicles. We show that passing the Turing test must be performed under conditions that are significantly different from those commonly used. We argue that passing such a test can be presumably much harder than the original. We consider a number of additional tests that can be used as some parts of the Turing test. In particular, we can mention such tests as practical test of driving skills, health test, test of prediction skills, security test, body test, coexistence test, aging test, trust test, overall performance test, no-harm test. In this paper, we pay special attention to the no-harm test. We consider an approach that is based on evolutionary machine learning. For the first study of the no-harm test, we have considered a relatively simple model of the natural environment and have proposed an algorithm for artificial evolution for the environment. The results of our experimental studies show that insufficiently justified implementation of autonomous vehicles can lead to unpredictable consequences.**

***Keywords*— Autonomous vehicles, evolutionary machine learning, no-harm test, Turing test.**

I. Introduction

It has been predicted that manufacturers will introduce automated vehicles into the market by 2025 (see e.g. [1]). It is assumed that by 2045 automated vehicles will have at least 80% share of the car market [2]. However, many fundamentally important questions remain unresolved. In particular, there are significant challenges in establishing uniform regulatory tests for the certification of autonomous vehicles [3]. Such certification should make it possible to evaluate the performance of autonomous vehicles with regard to road safety. Most researchers believe that such certification should be based on some analogue of the Turing test. However, views on the level of requirements for an analogue of the Turing test differ significantly. In some cases, it is assumed that it is enough to consider some certification framework based on the Turing test (see e.g. [4]). A number of studies argue for the need to develop additional tests and procedures. In particular, the replacement test has been proposed [5]. On the other hand, some researchers believe that the analogue of the Turing test for autonomous vehicles should be much easier than the original Turing test [6]. In particular, there are a number of reports that autonomous vehicles have passed the Turing test (see e.g. [7]). Moreover, some researchers believe that the Turing test can be passed in some relatively simple simulation settings [8-10].

In this paper, we do not consider the issue of the necessity or sufficiency of the Turing test. We show that passing the Turing test is usually considered under incorrect conditions. Passing the Turing test under the correct conditions will require significantly more effort from an artificial intelligence system. Moreover, the direct analogue of the Turing test for autonomous vehicles is significantly more difficult than the original Turing test.

II. Turing Test

Currently, there is no exact definition of intelligence for autonomous vehicles [6]. Accordingly, there is no clear understanding of the concept of an autonomous vehicle. There are a number of unresolved social dilemmas that are associated with the development of autonomous vehicle technologies. Even the question of the need for autonomous vehicles does not yet find a clear answer. In particular, it is mentioned in [11] that “the balance between the short-term benefits and long-term impacts of vehicle automation remains an open question”. At the same time, the need for autonomous vehicles is gaining increasing support. In particular, several countries allowed autonomous vehicles to be tested on ordinary roads.

The need for a global dialogue to establish a standard for artificial intelligence on our roads was expressed during the AI for Good Global Summit 2019. It is assumed that this standard will be some generally accepted analogue of the Turing test for autonomous vehicles. Perhaps this is just the future that rushes into the present too quickly. It is well known that when we make a wish, we should be careful. Our desire can be fulfilled. However, when this desire is fulfilled, we can understand that we did not get exactly what we wanted. However, we can better see the possible prospects of autonomous vehicles through the prism of the Turing test.

It is usually assumed that autonomous vehicles should be safer and more economical. Such vehicles can solve a number of problems. Among other things, we can mention

Manuscript received July 22, 2025.

A. A. Gorbenko is with the Ural Federal University, Ekaterinburg, Russian Federation (e-mail: gorbenko.aa@gmail.com).

V. Yu. Popov is with the Ural Federal University, Ekaterinburg, Russian Federation (e-mail: popovvvv@gmail.com).

accident, pollution, and traffic congestion. However, the usefulness of autonomous vehicles to solve some problems naturally implies that such vehicles are intelligent enough. Therefore, in order to prove the usefulness of autonomous vehicles, we need a proper vehicle intelligence test.

There are a number of different approaches that can make autonomous vehicles a reality. Some of these approaches imply the need for various adaptations of environments. In particular, encoded asphalt materials for the adaptation of pavements have been proposed [12]. Also, unmanned environments should be mentioned [13]. In general, we can consider various types of partially depopulated environments. For instance, we can provide fully autonomous environments [14]. Also, we can consider purposefully prestructured environments that are more or less not affected by individual and collective human activities. It is clear that attempts to adapt environments can be negatively perceived by society. However, such a reaction of society should not always be expected. There are improvements in the environment that can be considered as elements of additional comfort for human drivers and at the same time significantly simplify the provision of sustainable navigation for autonomous vehicles. In particular, the creation of global Wi-Fi coverage should be mentioned [15]. Many of the possible modifications to the environment should be considered as improvements that are aimed primarily at increasing road user safety. In particular, we can mention a more careful attitude to road surface markings and road fences, various improvements aimed at improving traffic safety in difficult weather conditions such as, for example, heated roads in cold regions, clarification of traffic rules. It is important to take into account that in many cases even relatively minor improvements can significantly reduce the demands on the level of intelligence of autonomous vehicles. For instance, moving into the on-coming lane to avoid an obstacle may require autonomous vehicles to use fairly specific human-level intelligence skills such as the ability to negotiate [16]. Not every human is able to demonstrate negotiability. Accordingly, the absence of such an ability should not affect the passage of the Turing test. At the same time, a minor refinement of traffic rules would make it possible to cope with such a maneuver based on a simple deterministic algorithm. Currently, the direction of research related to the modification of environments is underestimated. Despite the significant potential for solutions that can essentially reduce the requirements for autonomous vehicles and increase the sustainability of transport systems, there is only few number of investigations in this direction. Researchers focus on creating autonomous vehicles for ordinary roads. Frequently, it is assumed that autonomous vehicles should collaborate and coexist with humans safely and capably on the roads.

In recent years, the Turing test has been extensively studied in the context of autonomous vehicles [17-19]. The idea of creating a Turing test for autonomous vehicles for ordinary roads has been actively discussed in recent years. It should be noted that the Turing test is only a scientific experiment. The purpose of this test is merely to ascertain the agent's ability to think. If an agent passes the Turing test, we will get a scientific result and nothing more. Passing the Turing test is not a justification for the practical use of this agent. Also, passing the Turing test is not a recognition of the ability of this agent to function independently. It should be noted that the possible emergence of the Turing test for autonomous vehicles is a fundamentally important point for the development of technology of autonomous vehicles. Currently, the technology of autonomous vehicles is causing considerable doubt among researchers. However, most of the arguments voiced are related to the fact that autonomous vehicles may be unable to solve some problems. At the same time, the same problems are fundamentally difficult for humans. For instance, "they will sometimes have to choose between two evils, such as running over pedestrians or sacrificing themselves and their passenger to save the pedestrians" [20]. Emergence of the Turing test for autonomous vehicles will allow the technology to become a reality not because the vehicles can solve problems well, but because the vehicles can solve these problems like humans.

There are even some preliminary versions of such a test [21]. For instance, we can mention the ADA AV Turing Test. The ADA AV Turing Test is based on three principles to meet the burden of proof.

1. Prove artificial intelligence never engages in careless, dangerous or reckless driving behaviour.
2. Prove artificial intelligence remains aware, willing and able to avoid collisions at all times.
3. Prove artificial intelligence meets, or exceeds, the performance of a "competent & careful" human driver.

It is commonly accepted that we cannot just use the Turing test. It is quite natural. To obtain a driver's license, humans must not only demonstrate some knowledge but also driving skills. Moreover, humans must comply for health reasons. It seems that testing the health and driving skills is not implied by the spirit of the Turing test. Thus, the Turing test for autonomous vehicles should be a bit harder. However, usually it is assumed that passing such a test should be presumably much simpler than the original [16]. In particular, in some cases, it is assumed that the autonomous vehicle is enough to demonstrate only some limited form of intelligence [6]. It is possible that this is indeed so. However, in this case, it is preferable to obtain an explicit formulation of such limitations. The original Turing test has been proposed to answer the question "Can machines think?" Respectively, if we pose the question "Can machines drive?", what should be understood by the word "drive"? Also, it is necessary to find an explanation of how the autonomous vehicle with only some limited form of intelligence will solve the following problems.

If the vehicles will demonstrate the average behavior of human drivers, it is necessary to understand benefits of using such vehicles. Frequently, it is assumed that autonomous vehicles should at least meets the performance of careful human drivers. In general, an increase in the number of careful drivers should be welcome. However, in the case of autonomous vehicles, such an increase will happen without reinforcement by the corresponding changes in society. Such vehicles can upset the balance on the roads. A number of researchers have studied the implications of

this rebalancing for transport performance (see e.g. [1]). In particular, it has been shown that in some cases the introduction of autonomous vehicles can lead to a drop in productivity (see e.g. [22]). Therefore, an overall performance test is necessary. However, the issue of safety is significantly more important. An increase in the number of careful drivers can cause negative reactions of some other human drivers. In particular, this can lead to an increase in the number of provocations and, as a result, emergency situations.

Among other important skills, the vehicle must prove its ability to consider moral and ethical aspects when assessing the possibility of violating traffic rules such as speeding [23]. Studies of the interaction of a pedestrian and a robot have shown that participants changes their trajectory in the close proximity of the robot in 29 % of trials [24]. A similar situation is naturally expected for the vehicles. So, the vehicle should predict the intentions of other agents to achieve sufficient performance. Of course, the autonomous vehicle must prove a capability to avoid dangerous situations [20]. It seems obvious that the presence of such a capability implies a capability to predict the possibility of a dangerous situation. It is clear that we need some justification of the possibility of solving these problems by some limited form of intelligence.

Currently, there is no complete clarity as to what exactly can guarantee that the vehicles will not leave roads. In 2015, Twitter software robots have already demonstrated that they can be a threat, even though this behavior has not been programmed [25]. In addition, the actions of intruders can cause vehicles to leave roads [26]. Attempts to limit oneself to some form of road tests leave unanswered the question of the off-road behavior of the vehicles. The original Turing test has been proposed to answer the question “Can machines think?” Therefore, the purpose of the Turing test is to find out our opinion about an agent. If we want to allow the agent to coexist with humans, the agent must pass an additional test. Such a test should demonstrate the agent’s opinion of humans, the agent’s opinion on humans, and the agent’s opinion about humans. Moreover, the existence of an agent in time makes it necessary to pass another test. This test should determine the agent’s susceptibility to age-related changes. When passing the original Turing test, we implicitly assume that the artificial intelligence system is honest. However, the degree of success in passing the test does not depend on whether the system is fair or not. When we license an autonomous vehicle, we should trust that autonomous vehicle. Many artificial intelligence systems use Internet resources for training. For example, GPT-3. Using Internet resources for learning can lead to the creation of a deceptive artificial intelligence system. In particular, the creation of a deceptive artificial intelligence system may be due to the fact that the system has learned to use deception to its advantage. In other cases, the creation of a deceptive artificial intelligence system may be the result of the accumulation of faulty knowledge. Autonomous vehicles must pass a test that assesses the possible level of trust in the system.

Finally, attention should be paid to the existence of a natural approach to the problem of creating autonomous vehicles based on the use of a humanoid robot. If a humanoid robot can pass a well-defined original Turing test, then the robot can obtain a license to drive in the usual way and drive an ordinary car. In this case, the need to develop not only the Turing test for autonomous vehicles but also the autonomous vehicles themselves disappears. Usually, for general reasons, the task of creating such a humanoid is considered to be significantly more complicated. However, a comprehensive study of this issue has not been conducted. Even a comparatively superficial comparison of the vehicle and a humanoid driver allows us to formulate some important issues that should be verified by an analogue of the Turing test. The vehicle’s lack of body in itself raises a number of difficult issues. Some of these issues can be found to be consistent with the spirit of the Turing test. For instance, the influence of the absence of a body on the formation and development by an agent of a relation to the admissibility of risk, the value of another’s property, the value of his own car, the relation to the bodies of other agents, the value of his own life, the value of the lives of other agents, etc. However, some other issues imply the need for some additional verifications. Among others, we can mention the following issues. Under certain conditions, a human driver can solve a number of problems that are not directly related to driving. In particular, he can provide first aid. The ability of humans to drive does not mean their ability to live on the road. Virtual reality allows to study human behavior in a wide range of settings. The problem of misperception of egocentric distances in virtual environments is well known. A number of studies have shown that the ability to move significantly affects the perception of environments [27]. Thus, in the general case, the absence of a body can significantly affect the perceptions of the agent. Therefore, in addition to the Turing test, the vehicle must pass some kind of perception test. It is clear that additional verifications for these and previously considered issues can be included in the analogue of the Turing test. However, it seems that passing such a test should be much more difficult. In particular, at least some additional tests seem necessary. In particular, we can consider some important tests that should be established for autonomous vehicles. We assume that these tests must be passed by an autonomous vehicle but are presumably not implied by the spirit of the Turing test and, as a result, may not be provided by a direct analogue of the Turing test.

- **Practical test of driving skills.** The autonomous vehicle must demonstrate not only theoretical knowledge but also practical driving skills. It should be noted that the demonstration of practical skills can reveal some intellectual abilities that are difficult to verify during a theoretical test.
- **Health test.** Such a test concerns not only the technical condition of the autonomous vehicle. The original Turing test allows passing in relatively comfortable conditions. The autonomous vehicle must demonstrate high intelligence in relatively extreme conditions. In particular, the autonomous vehicle must demonstrate high intelligence when making quick decisions. In addition, the autonomous vehicle should not demonstrate a

significant decrease in intelligence during the working day.

- **Test of prediction skills.** The autonomous vehicle must be able to predict important events and intentions.
- **Security test.** The autonomous vehicle must demonstrate its ability to withstand the actions of intruders.
- **Body test.** The autonomous vehicle must demonstrate at least the absence of negative consequences that may be associated with a lack of body.
- **Coexistence test.** The autonomous vehicle must demonstrate an opinion that would allow us to conclude about the possibility of successful coexistence.
- **Aging test.** The autonomous vehicle must demonstrate a possibility of age-related changes.
- **Trust test.** An autonomous vehicle must prove that it is trustworthy.
- **Overall performance test.** The autonomous vehicle must justify that its appearance on the roads will not lead to a decrease in the productivity of the transport system.
- **No-harm test.** The autonomous vehicle must justify that its appearance on the roads will at least not cause harm.

III. NO-HARM TEST

The introduction of various robotic technologies, like the introduction of any new technology, can cause some significant changes in society, business and many other areas important to humans. Some changes should be considered as obviously desirable. Some changes require at least a more detailed analysis. Some changes may have significant negative consequences. Currently, there is no full understanding of the possible consequences. The study of the problem of unintended consequences of the introduction of robotic technologies is at the stage of forming research themes [28]. The potential consequences of autonomous transport should naturally be considered in the general context of robotic technologies. However, it is necessary to take into account the factor of acceptance of such technologies. It is well known that the acceptance of low complexity technologies is significantly higher [29]. Accordingly, the requirements for studying the possible consequences of the introduction of autonomous transport should be at least higher than average.

A large number of intelligent technologies are being extensively implemented in various fields. Many researchers believe that in the future intelligent agents will become the primary mode of human-computer interaction [30]. However, we currently do not have a full understanding of the possible consequences of the introduction of intelligent technologies. Moreover, there are no generally accepted views on the assessment of such consequences. In particular, some researchers argue for the possibility of foreseeable and anticipated harm in the context of transformative service systems [31-33]. Some other researchers argue that no intended harms exist in the context of transformative service systems [34]. It should be noted that the problems of harmful robotic actions attract significant attention of researchers [35-37]. Some researchers believe that people should not trust robots because of the potential harm that can come from interacting with robots [38]. In some cases, avoiding harmful actions is seen as a duty of robots [39]. However, there are also completely different views on this issue. In particular, we can mention a new socio-technological ethical framework of augmented utilitarianism that has been developing extensively in recent years [40-43]. Augmented utilitarianism does not represent a normative theory [44]. Augmented utilitarianism suggests that developers should avoid solving difficult moral problems [45]. Moral boundaries should be programmed by users [45]. Thus, within the framework of the augmented utilitarianism, the moral character of a robot strictly depends on the current mood of the user. Taking into account the mechanisms triggering robot abuse [46-48], augmented utilitarianism allows for the emergence of harmful robots as manifestations of a mocking attitude or hostility. It should be noted that some people abuse robots, believing that such actions are morally acceptable [49].

The various potential harmful consequences of the introduction of autonomous vehicles represent too broad an area of research. In this paper we have consider only some aspects of road safety. Improved road safety is typically seen as one of the main benefits of introducing autonomous vehicles [50,51]. When pointing out the possibility of achieving such benefit, researchers sometimes cite news and the results of mathematical modeling (see e.g. [51]). It is often assumed that the higher skills of autonomous drivers should lead to improved road safety. However, mathematical modeling results show that demonstrating the skills would take approximately 400 years [52]. Moreover, the analysis of [52] shows that some aspects cannot be demonstrated. In this paper we do not consider the problem of demonstrating skills. We consider the relationship between skill level and road safety in terms of road accidents.

Investigations of [53] indicate microsimulations as the main approach to traffic modeling. It should be noted that simulations are extensively used to solve various autonomous driving problems (see e.g. [54,55]). In particular, practical research of safety problems is significantly complicated by the relative rarity of road accidents and the undesirability of their reproduction in the real world. Random testing is quite acceptable for initial investigations (see e.g. [56]). Our approach is based on evolutionary machine learning methods [57]. For the first study of the problem of non-harm, we have considered a relatively simple environment. We are considering a model of artificial evolution for the environment. In particular, we have used an evolutionary generative model that is based on descriptive representations [57]. It is assumed that evolutionary computation is supported by self-organising maps [57] that are used to categorize individuals and manage memory. It should be noted that individuals are not completely determined by the general evolutionary generative model.

The general evolutionary generative model determines only the values of the main characteristics of individuals. Each individual has its own behavior. However, the behavior of each individual is based only on the model of simple Darwinian evolution. We deliberately avoid complex behavior patterns to reduce their impact on the performance of the general model. It is assumed that all individuals move along the ring road in the same direction.

Each individual starts moving from a parking lot located at a random point on the road. The length of the route is three

circles. The motion ends at the parking lot. The motion resumes immediately after returning to the parking lot. The road has four lanes and a shoulder on the right. Driving on the shoulder is considered a violation. The system randomly generates a parking requirement on the shoulder.

Initially, all individuals have random behavior settings and do not have combos. The behavior of individuals is determined by the basic parameters, aggressiveness, accuracy, discipline, attentiveness, ingenuity, conservatism, altruism. In addition, the behavior of individuals is determined by their preference for lanes depending on traffic. Aggressiveness determines the level of danger of maneuvers and the frequency of lane changes. Accuracy determines the accuracy of maneuvers. Discipline determines the tendency to break rules. Attention determines the visible area of the road. Ingenuity determines the complexity of the combos and their number. Conservativeness determines the tendency to change behavior parameters and change combos.

As a fundamental basis for simulations of changes of parameter settings, we consider BORCGA-BOPSO hybrid genetic algorithm that has been proposed in the paper [58]. In particular, replacing the objective functions with appropriate inverse values is used to reduce the multiobjective optimization problem to a biobjective minimization problem. The generation of new combos is based on a simple artificial evolution model that uses the complexity of the combo as an input parameter. The complexity of the combo is adjusted by a recurrent neural network based on the success of previously generated combos. The choice between a changing parameter settings and a creating combos is determined by adversarial neural networks [59]. The functioning of neural networks is supported by the model of reinforcement learning that has been considered in the paper [60]. To introduce randomness, we consider two linear congruential generators with the recursive formulas

$$X[i+1] = (16807X[i]) \bmod (2^{31}-1),$$

$$X[i+1] = (314159269X[i]) \bmod (2^{31}-1)$$

(see e.g. [61,62]). To generate random numbers, we use a shuffle F of the linear congruential generators with the steering Fibonacci word (see e.g. [63-65]). We consider random functions in the form

$$f(x) = \sum_{k=0}^{r} a[k]x^k,$$

where it is assumed that $a[k], 0 \le k \le r,$ are random numbers whose values are determined by the generator *F*, *r* is a random number which is determined by the lagged Fibonacci generator with the recursive formula

$$X[i+31] = (X[i+28] + X[i]) \bmod (2^{32})$$

(see e.g. [66]). The pseudo-random number generator XorShift has been used for seed generation [67].

Altruism defines the tendency to take into account the interests of other individuals. Each parameter can take values from 1 to 100. The success of the next passage of the route for each individual is determined by the time and number of dangerous situations. One of the values is basic. The second value should be within the acceptable range. If the discipline is more than 50 then the number of dangerous situations is basic. If the aggressiveness is more than 50 or the discipline is less than 50 then the time is basic. Successful completions are the basis for creating combos. Insufficient success requires changing parameter settings or creating combos. The level of change depends on the value of conservatism. If the conservativeness value is less than 30, then the settings change even after successful runs. After 5000 generations of evolution, the population is divided into three groups depending on the value of the discipline, disciplined individuals (10%), normal individuals (20%), undisciplined individuals (70%). We remove some undisciplined individuals from the population and add the same number of disciplined or normal artificial individuals. We obtain artificial individuals for population update by cloning natural individuals. Because the results of our simulations depend significantly on a number of random choices, we consider only the average values over 50 restarts of evolution. We have considered changes in the natural population by adding 10%, 20%, 30%, 40%, 50%, 60% disciplined and normal individuals. The natural evolutionary background is given in Tab. 1, 2. We consider the results of evolution as the dependence of the change in the number of road accidents on the number of generations in Tab. 1. In particular, we consider three main groups of road accidents, vehicle collisions (C), violations of rules (V), dangerous maneuvers without formally violating the rules (D). In Tab. 2, we consider the dependence of the change in the population structure (disciplined individuals (X), normal individuals (Y), undisciplined individuals (Z)) on the number of generations.

TABLE I. ACCIDENT RATES FOR THE NATURAL POPULATION

G	C	V	D
100	100%	100%	100%
200	99.61%	98.83%	99.38%
1000	98.73%	97.28%	99.12%
2000	97.28%	96.11%	93.19%

TABLE II. EVOLUTION OF THE NATURAL POPULATION

G	X	Y	Z
100	10.12%	21.04%	68.84%
200	10.17%	22.85%	66.98%
1000	11.13%	34.27%	54.60%
2000	12.03%	42.19%	45.78%

In most cases, we have obtained extremely unstable results. Relatively consistent results have been obtained only for a large number of normal individuals (see Tab. 3, 4) and a small number of disciplined individuals (see Tab. 5, 6).

TABLE III. ACCIDENT RATES FOR A NORMAL ADDITION

G	C	V	D
100	116.82%	27.22%	144.15%
200	76.19%	23.56%	87.94%
1000	7.21%	11.47%	22.78%
2000	8.33%	12.77%	24.57%

TABLE IV. EVOLUTION WITH A NORMAL ADDITION

G	X	Y	Z
0	10%	70%	20%
100	9.34%	67.81%	22.85%
200	8.96%	61.19%	29.85%
1000	5.38%	83.91%	10.71%
2000	5.44%	84.14%	10.42%

TABLE V. ACCIDENT RATES FOR A DISCIPLINED ADDITION

G	C	V	D
100	74.25%	66.77%	378.12%
200	88.66%	76.92%	325.86%
1000	126.33%	96.43%	239.24%
2000	142.11%	112.21%	242.97%

TABLE VI. EVOLUTION WITH A DISCIPLINED ADDITION

G	X	Y	Z
0	30%	20%	50%
100	32.17%	9.43%	58.40%
200	21.72%	23.52%	54.76%
1000	19.18%	27.71%	53.11%
2000	18.32%	29.94%	51.74%

A significant increase in normal individuals leads to some deterioration in safety. However, over time the situation improves significantly even in comparison with the initial. A small increase in the number of disciplined individuals has an obvious negative effect. For other cases, unstable results have been obtained under our conditions.

For the case of the significant increase in normal individuals, we considered an additional option. The ability to unlearn is a well-known adaptation mechanism [18,68]. In general, there are several different approaches to unlearning. Individuals have acquired the skill not only to create and complicate combos, but also to simplify and forget combos. In particular, it is assumed that the increase in dangerous situations requires the abandonment of the use of the most complex well-learned combos. The simplest actions or simplified versions of such combos should be used. The increase in dangerous situations may be a manifestation of two related facts.

- The traffic situation has changed significantly.
- The usual proven tactics are situationally dependent.

In such conditions, the driver should abandon the use of proven tricks and act as simply as possible. A stable change in the traffic situation should lead to a complete rejection of proven combos or at least their radical revision. We have reproduced the experiments for the normal addition under the same conditions but with the unlearning option. The unlearning process is regulated by a generative adversarial neural network [59]. Typically, unlearning is used to discard obsolete, redundant, and incorrect knowledge. Usually it is assumed that there are sufficiently clear criteria for identifying such knowledge. In our case, both combo generation and unlearning formally pursue a common purpose, improving driver performance. There is no clear evidence for the correct choice between generation and unlearning. So, proper management of the unlearning process requires taking into account a large number of significantly different parameters. In particular, we can mention such parameters as driver behavior, driver characteristics, current set of combos, specific changes in road traffic, specifics of the current dangerous situation. It is clear that the development of approaches to optimal unlearning requires a separate study. So, for initial investigations, we consider a generative adversarial neural network as a simple black box advisor. It is assumed that only new individuals have the skill of unlearning. The results of the experiments are given in Tab. 7, 8.

TABLE VII. ACCIDENT RATES WITH UNLEARNING

G	C	V	D
100	98.31%	32.85%	119.39%
200	43.45%	21.13%	76.18%
1000	6.17%	10.23%	19.11%
2000	5.68%	9.48%	17.88%

TABLE VIII. EVOLUTION WITH UNLEARNING

G	X	Y	Z
0	10%	70%	20%
100	9.88%	69.32%	20.8%
200	9.23%	67.54%	23.23%
1000	7.91%	85.42%	6.67%
2000	7.74%	87.28%	4.98%

The results of the experiments demonstrate a significant reduction in the number of collisions and dangerous situations. The number of disciplined individuals decreases more slowly.

IV. CONCLUSION

In this paper we have argued that the direct analogue of the Turing test for autonomous vehicles is significantly more hard than the original Turing test. We have proposed a number of additional tests that could be used to license autonomous vehicles. Our experimental results reflect only initial study of the no-harm problem. More needs to be done to finalize a “Hippocratic Oath” for autonomous vehicles. Nevertheless, we have obtained sufficient results to state that the transition to mixed traffic can lead to a significant decrease in safety. In particular, replacing undisciplined drivers with disciplined ones generally does not lead to an increase in road safety. Accordingly, it is necessary to find special strategies to modify the qualitative structure of the driver population. Currently, the introduction of autonomous vehicles requires either a complete transition to autonomous drivers or maintaining the qualitative structure of the population. However, in some cases, a significant reduction in the number of road accidents and some improvement in

the population structure are possible. So, the proposed model of evolutionary machine learning can be used for further research. In particular, the model can be used to find specific strategies for modifying the qualitative structure of the driver population.

REFERENCES

[1] J. E. Park, W. Byun, Y. Kim, H. Ahn, D. and K. Shin, "The impact of automated vehicles on traffic flow and road capacity on urban road networks," JAT, vol. 2021, pp. 1–10, November 2021.

[2] T. Litman, "Autonomous vehicle implementation prediction: implications for transport planning," Victoria Transport Policy Institute, Victoria, Canada, 2015.

[3] D. Alipour and H. Dia, "A review of AI ethical and moral considerations in road transport and vehicle automation," in Handbook on Artificial Intelligence and Transport, H. Dia Ed. Cheltenham: Edward Elgar Publishing, 2023, pp. 534–566.

[4] H. Dia, R. Tay, R. Kowalczyk, S. Bagloee, E. Vlahogianni, and A. Song, "Artificial intelligence tests for certification of autonomous vehicles," Academia Letters, 2021.

[5] J. Sifakis and D. Harel, "Trustworthy autonomous system development," ACM T. Embed. Comput. S., vol. 22, pp. 1–24, April 2023.

[6] L. Li, W.-L. Huang, Y. Liu, N.-N. Zheng, and F.-Y. Wang, "Intelligence testing for autonomous vehicles: a new approach," IEEE Trans. Intell. Veh., vol. 1, pp. 158–166, June 2016.

[7] E. Cascetta, A. Cartenì, and L. Di Francesco, "Do autonomous vehicles drive like humans? A Turing approach and an application to SAE automation Level 2 cars," Transp. Res. Part. C Emerg. Technol., vol. 134, 103499, January 2022.

[8] N. J. Cowan and R. J. Full, "Swimming in the "Matrix": VR fish pass the Turing test using a simple control law for collective behavior", Sci. Robot., vol. 10, April 2025.

[9] L. Li, M. Nagy, G. Amichay, R. Wu, W. Wang, O. Deussen, D. Rus, and I. D. Couzin, "Reverse engineering the control law for schooling in zebrafish using virtual reality," Sci. Robot., vol. 10, April 2025.

[10] Y. Zhang, P. Hang, C. Huang, C. Lv, "Human-Like Interactive Behavior Generation for Autonomous Vehicles: A Bayesian Game-Theoretic Approach with Turing Test," Adv. Intell. Syst., vol. 4, January 2022.

[11] D. Milakis, B. van Arem, and B. van Wee, "Policy and society related implications of automated driving: A review of literature and directions for future research," J. Intell. Transp. Syst., vol. 21, pp. 324–348, 2017.

[12] F. Moreno-Navarro, G. R. Iglesias, and M. C. Rubio-Gamez, "Encoded asphalt materials for the guidance of autonomous vehicles," Autom. Constr., vol. 99, pp. 109–113, March 2019.

[13] H. Sun, and P. K. Venuvinod, "The human side of holonic manufacturing systems," Technovation, vol. 21, pp. 353–360, June 2001.

[14] X. Zhang, W. Liu, S. T. Waller, and Y. Yin, "Modelling and managing the integrated morning-evening commuting and parking patterns under the fully autonomous vehicle environment," Transport. Res. B., vol. 128, pp. 380–407, October 2019.

[15] E. I. Adegoke, J. Zidane, E. Kampert, C. R. Ford, S. A. Birrell, and M. D. Higgins, "Infrastructure Wi-Fi for connected autonomous vehicle positioning: A review of the state-of-the-art," Vehicular Communications, vol. 20, 100185, December 2019.

[16] N. Chater, J. Misyak, D. Watson, N. Griffiths, and A. Mouzakitis, "Negotiating the Traffic: Can Cognitive Science Help Make Autonomous Vehicles a Reality?," Trends. Cogn. Sci., vol. 22, pp. 93–95, February 2018.

[17] P. Bhavsar, I. Safro, B. Nidhal, P. Robi, D. Dera, P. Dutta, O. Aminul, "Chapter 13 – Artificial intelligence in transportation data analytics," in Data Analytics for Intelligent Transportation Systems. Elsevier, 2024, pp. 337–382.

[18] A. Bindayel, H. Elsayed, and M. U. Khan, "AI and self-reflection," LNCS, vol. 15819, pp. 289–302, May 2025.

[19] H. Lu, M. Zhu, and H. Yang, "Human-like driving technology for autonomous electric vehicles," Nat. Rev. Electr. Eng., vol. 2, pp. 218–219, March 2025.

[20] J.-F. Bonnefon, A. Shariff, and I. Rahwan, "The social dilemma of autonomous vehicles," Science, vol. 352, pp. 1573–1576, Jun 2016.

[21] S. F. Kalik and D. V. Prokhorov, "Automotive Turing Test," in Proceedings of the 2007 Workshop on Performance Metrics for Intelligent Systems. New York: ACM Press, August 2007, pp. 152–158.

[22] J. Vander Werf, S. E. Shladover, M. A. Miller, and N. Kourjanskaia, "Effects of adaptive cruise control systems on highway traffic flow capacity," Transp. Res. Rec., vol. 1800, pp. 78–84, January 2002.

[23] P. A. Hancock, I. Nourbakhsh, and J. Stewart, "On the future of transportation in an era of automated and autonomous vehicles," PNAS USA, vol. 116, pp. 7684–7691, January 2019.

[24] C. Vassallo, A.-H. Olivier, P. Soueres, A. Cretual, O. Stasse, and J. Pettre, "How do walkers avoid a mobile robot crossing their way?," Gait & Posture, vol. 51, pp. 97–103, January 2017.

[25] W. Hartzog, "Unfair and Deceptive Robots," Maryland Law Review, vol. 74, pp. 785–829, 2015.

[26] X. Liu, R. Jiang, H. Wang, and S. S. Ge, "Filter-based secure dynamic pose estimation for autonomous vehicles," IEEE Sens. J., vol. 19, pp. 6298–6308, April 2019.

[27] P. Maruhn, S. Schneider, and K. Bengler, "Measuring egocentric distance perception in virtual reality: influence of methodologies, locomotion and translation gains," PLoS ONE, vol. 14, e0224651, October 2019.

[28] N. Heirati, V. Pitardi, J. Wirtz, C. Jayawardhena, W. Kunz, and S. Paluch, "Unintended consequences of service robots – Recent progress and future research directions," J. Bus. Res., vol. 194, 115366, May 2025.

[29] F. F. M. Ferreira, P. P. Barbosa, L. C. Cesario, F. L. Lizarelli, and G. H. S. Mendes, "Customer acceptance of service robots: Marketing outcomes and service complexity," Serv. Mark. Q., August 2025.

[30] P. Bornet and J. Wirtz, Agentic Artificial Intelligence: Harnessing AI Agents to Reinvent Business, Work, and Life. Singapore: World Scientific Publishing Co. Pte. Ltd., 2025.

[31] C. P. Blocker, B. Davis, and L. Anderson, "Unintended consequences in transformative service research: Helping without harming," J. Serv. Res., vol. 25, pp. 3–8, December 2022.

[32] J. Finsterwalder and V. G. Kuppelwieser, "Intentionality and transformative services: Wellbeing co-creation and spill-over effects," J. Retail. Consum. Serv., vol. 52, 101922, January 2020.

[33] W. J. FitzPatrick, "The doctrine of double effect: Intention and permissibility," Philos. Compass, vol. 7, pp. 183–196, March 2012.

[34] M. J. Polonsky, V. Weber, L. Ozanne, and N. Robertson, "A framework of foreseen and unforeseen harms in transformative service systems," J. Serv. Res., vol. 0, pp. 1–19, September 2024.

[35] A. Robey, Z. Ravichandran, V. Kumar, H. Hassani, and G. J. Pappas, "Jailbreaking LLM-controlled robots," in 2025 IEEE International Conference on Robotics and Automation. Piscataway: IEEE, September 2025.

[36] A. Hilliard, E. Kazim, and S. Ledain, "Are the robots taking over? On AI and perceived existential risk," AI Ethics, vol. 5, pp. 2929–2942, June 2025.

[37] R. Zhang, H. Li, H. Meng, J. Zhan, H. Gan, Y.-C. Lee, "The dark side of AI companionship: A taxonomy of harmful algorithmic behaviors in human-AI relationships," in CHI '25: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems. New York: Association for Computing Machinery, May 2025, pp. 1–17.

[38] G. M. Gomez, "Should we trust social robots? Trust without trustworthiness in human-robot interaction," Philos. Technol., vol. 38, February 2025.

[39] J. Haltaufderheide, S. Pfisterer-Heise, D. Pieper, and R. Ranisch, "The ethical landscape of robot-assisted surgery: a systematic review," J. Robotic. Surg., vol. 19, March 2025.

[40] C. Gros, M. Martens, N.-M. Aliman, L. Kester, and P. Werkhoven, "Ethical frameworks for automated vehicles: a systematic analysis and design," AI Ethics, August 2025.

[41] C. Gros, L. Kester, M. Martens, P. Werkhoven, "Addressing ethical challenges in automated vehicles: bridging the gap with hybrid AI and augmented utilitarianism," AI Ethics, vol. 5, pp. 2757–2770, June 2025.

[42] N.-M. Aliman and L. Kester, "Requisite variety in ethical utility functions for AI value alignment," CEUR Workshop Proceedings, vol. 2419, 2019.

[43] N.-M. Aliman, L. Kester, P. Werkhoven, "XR for Augmented Utilitarianism," in 2019 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR). Piscataway: IEEE, December 2019.

[44] N.-M. Aliman and L. Kester, "Augmented utilitarianism for AGI safety," in Artificial General Intelligence. Cham: Springer, July 2019, pp. 11–21.

[45] N.-M. Aliman and L. Kester, "Moral programming," in Moral design and technology. Wageningen: Wageningen Academic Publishers, January 2022, pp. 63–80.

[46] Y. Chen, J. Cao, S. Liu, M. Huang, F. Wan, and C. W. Chen, "Friend or foe? An empirical study on the mechanisms triggering robot abuse," J. Hosp. Tour. Manag., vol. 64, 101318, September 2025.
[47] M. Shiomi, K. Sakai, T. Funayama, T. Minato, and H. Ishiguro, "Robot harassment: Inappropriate behaviors against an android robot during three-year exhibition," in 2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI). Piscataway: IEEE, March 2025.
[48] A. Rezzani, M. Menendez-Blanco, B. J. Bushman, and A. De Angeli, "What is robot abuse? A sociotechnical definition," Behav. Inform. Technol., February 2025.
[49] S. Yamada, T. Kanda, and K. Arimitsu, "Development and validation of moral concern for robot scale (MCRS)," ACM T. Hum.-Robot Inte., vol. 14, pp. 1–39, July 2025.
[50] A. Uzzaman, M. I. Adam, S. Alam, and P. Basak, "Review on the safety and sustainability of autonomous vehicles: Challenges and future directions," Control Syst. Optim. Lett., vol. 3, pp. 103–109, January 2025.
[51] M. Naiseh, J. Clark, T. Akarsu, Y. Hanoch, M. Brito, M. Wald, T. Webster, and P. Shukla, "Trust, risk perception, and intention to use autonomous vehicles: an interdisciplinary bibliometric review," AI & Soc., vol. 40, pp. 1091–1111, February 2025.
[52] N. Kalra and S. M. Paddock, "Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability?," Transp. Res. A Policy Pract., vol. 94, pp. 182–193, December 2016.
[53] J. E. Park, W. Byun, Y. Kim, H. Ahn, and D. K. Shin, "The impact of automated vehicles on traffic flow and road capacity on urban road networks," J. Adv. Transp., vol. 2021, 8404951, November 2021.
[54] S. He, X. Guo, F. Ding, Y. Qi, and T. Chen, "Freeway traffic speed estimation of mixed traffic using data from connected and autonomous vehicles with a low penetration rate," J. Adv. Transp., vol. 2020, 1361583, June 2020.
[55] X. Hu, S. Li, T. Huang, B. Tang, R. Huai, and L. Chen, "How simulation helps autonomous driving: a survey of Sim2real, digital twins, and parallel intelligence," IEEE Trans. Intell. Veh., vol. 9, pp. 593–612, January 2024.
[56] H. Zhang, J. Sun, and Y. Tian, "Accelerated safety testing for highly automated vehicles: Application and capability comparison of surrogate models," IEEE Trans. Intell. Veh., vol. 9, pp. 2409–418, January 2024.
[57] W. Banzhaf, P. Machado, and M. Zhang, Handbook of evolutionary machine learning. Singapore: Springer, 2024.
[58] A. S. Akopov and L. A. Beklaryan, "Traffic improvement in Manhattan road networks with the use of parallel hybrid biobjective genetic algorithm," IEEE Access, vol. 12, pp. 19532–19552, February 2024.
[59] Y. Li, F. Bai, C. Lyu, X. Qu, Y. Liu, "A systematic review of generative adversarial networks for traffic state prediction: Overview, taxonomy, and future prospects," Inf. Fusion, vol. 117, May 2025.
[60] S. Yoo, H. Kim, and J. Lee, "Combining reinforcement learning with genetic algorithm for many-to-many route optimization of autonomous vehicles," IEEE Access, vol. 12, pp.26931–26942, February 2024.
[61] P. Bratley, B. Fox, and L. Schrage, A Guide to Simulation. New York: Springer Verlag, 1983.
[62] E. Taillard, "Benchmarks for basic scheduling problems," Eur. J. Oper. Res., vol. 64, pp. 278–285, January 1993.
[63] L. Balkova, M. Bucci, A. De Luca, and S. Puzynina, "Infinite words with well distributed occurrences," in Combinatorics on Words. Berlin, Heidelberg: Springer, August 2013, pp. 46–57.
[64] E. Charlier, T. Kamae, S. Puzynina, and L. Zamboni, "Infinite self-shuffling words," J. Comb. Theory, Ser. A, vol. 128, pp. 1–40, November 2014.
[65] L.-S. Guimond, J. Patera, and J. Patera, "Combining random number generators using cut-and-project sequences," Czech. J. Phys., vol. 51, pp. 305–311, April 2001.
[66] M. Matsumoto, I. Wada, A. Kuramoto, and H. Ashihara, "Common defects in initialization of pseudorandom number generators," ACM Trans. Model. Comput. Simul., vol. 17, pp. 15-1–15-20, September 2007.
[67] G. Marsaglia, "Random number generators," J. Mod. Appl. Stat. Methods, vol. 2, pp. 2–13, January 2003.
[68] Y. Cao, J. Yang, "Towards making systems forget with machine unlearning," in 2015 IEEE Symposium on Security and Privacy. Piscataway: IEEE, May 2015, pp. 463–480.